Synthese Library 422

Studies in Epistemology, Logic, Methodology, and Philosophy of Science

José L. Falguera
Concha Martínez-Vidal *Editors*

Abstract Objects

For and Against

Springer

Synthese Library

Studies in Epistemology, Logic, Methodology,
and Philosophy of Science

Volume 422

The aim of *Synthese Library* is to provide a forum for the best current work in the methodology and philosophy of science and in epistemology. A wide variety of different approaches have traditionally been represented in the Library, and every effort is made to maintain this variety, not for its own sake, but because we believe that there are many fruitful and illuminating approaches to the philosophy of science and related disciplines.

Special attention is paid to methodological studies which illustrate the interplay of empirical and philosophical viewpoints and to contributions to the formal (logical, set-theoretical, mathematical, information-theoretical, decision-theoretical, etc.) methodology of empirical sciences. Likewise, the applications of logical methods to epistemology as well as philosophically and methodologically relevant studies in logic are strongly encouraged. The emphasis on logic will be tempered by interest in the psychological, historical, and sociological aspects of science.

Besides monographs *Synthese Library* publishes thematically unified anthologies and edited volumes with a well-defined topical focus inside the aim and scope of the book series. The contributions in the volumes are expected to be focused and structurally organized in accordance with the central theme(s), and should be tied together by an extensive editorial introduction or set of introductions if the volume is divided into parts. An extensive bibliography and index are mandatory.

More information about this series at http://www.springer.com/series/6607

José L. Falguera • Concha Martínez-Vidal
Editors

Abstract Objects

For and Against

Springer

Editors
José L. Falguera
Department of Philosophy
and Anthropology
University of Santiago de Compostela
Santiago de Compostela, Spain

Concha Martínez-Vidal
Department of Philosophy
and Anthropology
University of Santiago de Compostela
Santiago de Compostela, Spain

Synthese Library
ISBN 978-3-030-38241-4 ISBN 978-3-030-38242-1 (eBook)
https://doi.org/10.1007/978-3-030-38242-1

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The idea for this volume came from a series of workshops of the research group EPISTEME organized at the University of Santiago de Compostela, thanks to the grants awarded by FEDER together with the Spanish Ministry of Economy, Industry and Competitiveness (FFI2013-41415-P) and the Spanish Ministry of Science, Innovation and Universities (State Research Agency) (FFI2017-82534-P). Some of the contributed chapters of this volume are versions of lectures given in those workshops, while others are from leading authors invited to complete the volume. We, the editors of this volume, are grateful to all the authors who agreed to participate, thus making this project possible. We are also grateful to the editor of the series, Otávio Bueno, for his support all along.

The projects financed address central debates involving abstract objects. The relevance of the issue for our intellectual lives is not difficult to note. It is easy to see the ubiquity of expressions by means of which (putative) abstract entities are referred to in human discourses, for example, when using mathematical expressions (such as numerals or those that designate sets); when speaking about the content of sentences and beliefs (propositions); when doing so about the contents of general terms (concepts), about types of sentences and of expressions, or about possible worlds; for some, when talking about properties and relations (as universals); and, also, when talking about written works (i.e., novels, tales, etc.), musical works, institutions, and even fictional characters. This phenomenon is well known and frequently mentioned in the literature (see Moltmann 2013 and in this volume for a discussion of this phenomenon in terms of her core versus periphery distinction).

Given such ubiquitous use of expressions that purportedly refer to abstract objects, we think that it is relevant to attend to the controversy between those who want to advocate the existence of abstract objects and those who stand against them. This volume seeks to collect some controversies involving this issue in various philosophical fields. Thus, the object of the volume is neither to discuss generically whether abstract entities exist nor to focus on a mere specific area. Rather, the purpose is to bring together considerations related to different philosophical domains.

For those discussing abstract objects, it is important to clarify what they would be if they existed. The answer is not easy, and there are different ways to elucidate this question. (For a discussion of the issue, see "Introduction" by Plebani in this volume). Let us assume, for the sake of brevity, that they are not concrete entities (nor mental entities, if conceived as clearly different from concrete ones). By doing so, we are adopting a negative way to answer (for this and other ways to elucidate the notion of "abstract object," see Lewis (1986) and Rosen (2008)); however, this negative characterization does not include considerations about their possible a-causality or a-temporality. We reject to add a-causality because to say that a putative abstract object causes anything is problematic. For example, if a tale, understood as a complex content (not as its mere token copy), is a complex abstract entity—take for instance, *The History of Tom Thumb*—or if a character, as "Tom Thumb," is an individual abstract object, then it makes sense to say that the story or the character depicted in it causes behaviors (ways of behaving and even specific behaviors) in many people (especially children). And we reject to add a-temporality because if we assume that there are *abstract artifacts*, in Thomasson's sense (see her 1999), they seem to have a temporal dimension. Consider a character, like "Sancho Panza," or an institution, like a university; they are abstract, and it is not easy to assume that they do not have a beginning and surely an end, because they are created and they could disappear (e.g., if no copy of the novel *The Ingenious Gentleman Don Quixote of La Mancha* survived a catastrophe or if a university stops having any activity and closes). Anyway, by saying that an abstract object is not a concrete one, we are not solving the problem of characterizing abstract objects; rather, we are taking this simple and, maybe, not very problematic answer as a departing point.

Taking the abovementioned negative characterization of abstract objects as a working hypothesis, we find that the most typical cases of putative abstract objects are mathematical entities. Mathematical expressions and their alleged references have to do with those areas of knowledge that we consider the most excellent—the mathematical sciences—and those of the empirical sciences that use these expressions. It is not surprising that a very significant part of the philosophical literature on abstract objects—for and against—has been concerned with such putative entities. The seek for an understanding of mathematics, and of their contribution to knowledge in the empirical sciences, has prompted the question of the existence of mathematical entities. In case they do not, the issue is how to understand those mathematical statements that we usually consider correct, as well as how to understand those mathematical expressions that contribute to assertions that we identify as true.

Besides, philosophical elucidation in different areas (especially in the philosophy of language and the philosophy of mind) usually postulates entities such as propositions, concepts, or possible worlds, among others. In doing so, it seems that considering them is essential to provide satisfactory explications (in the Carnapian sense) on very different matters.

The mere consideration of the cases related to mathematics, as well as those related to philosophical elucidation, justifies that philosophy itself is concerned with clarifying whether the supposed entities referred to by such expressions and discourses really exist (as nonspatial entities that are neither mental nor of a supposed spiritual nature). At the end of the day, if they exist, they are mysterious in principle since our cognitive access to them is reached neither by perception nor by the empirical testing of hypotheses about such entities (about their existence, their nature); that is, they are not accessible via the common avenues of cognitive access to things in the material or concrete world.

In any case, among those who defend that they exist—known as realists about such objects and often as platonists—proposals have been put forward to establish why they are unavoidable. Among the most generic, we find, for example, those that appeal to the indispensability, especially the indispensability of mathematical entities for well-confirmed scientific theories (see Quine 1948/1953, 1969; Colyvan 2001; also Baron, in this volume; Martínez-Vidal and Rivas for current state of the debate and, against, for example, Leng 2002), or those that appeal to their existence according to some principle, such as *Zalta's comprehension axiom* for abstract objects (see Zalta 1983, 1989, and in this volume)—which incorporates any abstract object for which there is a (mental or linguistic) representation with a unique content, including absurd or contradictory objects as *the round square*—or Balaguer's *plenitudinous platonism* (see Balaguer 1998 and in this volume). Balaguer argues that his plenitudinous platonism is the best way to advocate for abstract objects, though he does not endorse it: in fact, he maintains a neutral position about the existence of mathematical objects because he takes there are also good arguments against their existence and there is no way to decide (see also M. Leng in this volume for other ways of arguing for a similar position). Also, neo-Fregeans appeal to abstraction principles, many of which share a similar structure with the famous *Hume's principle* (see Wright 1983 and Hale 1987, both based in Frege 1884).

Some people who adopt a realist or platonist approach on this issue advocate that abstract objects are individuals, others defend that the only abstract objects are structures (interrelated properties and relations), while some others assume that both (individuals and structures) are citizens of the universe of abstract objects. As a result, defenders of abstract objects have developed different views about their nature. Platonists need to provide an appropriate account of our epistemic access to such (in principle) mysterious entities; among these kinds of accounts, we find (i) naturalistic proposals that try to explain how abstract objects obtain from empirical entities; (ii) views according to which they are grounded in empirical entities; (iii) those who contend and grasp them thanks to some kind of especial perception of abstract entities; or (iv) that they result from abstraction processes; or (v) thanks to our grasping (as a mental concept) what exists as an (abstract) intersubjective concept, or (vi) by capturing abstract objects through consistent comprehensive theories, or (vii) by considering that abstract objects are created or constituted by human beings (or, in general, by intelligent beings).

Frequently, people who believe in the existence of abstract objects, of any kind, are assuming that we—human beings—clearly know about some concrete objects. The point here is that they not only assume the existence of some concrete objects but also that we know some existing concrete objects—among them some more conjecturable as so-called by philosophers of science "theoretical entities"—and that in order to know these concrete entities, we indispensably need abstract objects, like mathematical objects, for example. This idea underlies the Quinean ontological and meta-ontological approach (see Quine 1953, 1969), i.e., the predominant ontological and meta-ontological paradigm all along the second half of the twentieth century. Nevertheless, according to some neo-Carnapian meta-ontological approaches, it is possible to reject that we can get clear and definitive knowledge about which concrete objects there are in the sense of hard metaphysics. In consequence, they contend, for example, that there must exist some concrete entities though it is not possible for us to discover which. Furthermore, some go beyond to argue that while what we establish about concrete objects is conjectural, what we establish about abstract objects is not; that would be so because abstract objects are created or constituted (as abstract artifacts), while concrete objects are not (see de Donato-Rodríguez and Falguera, Chap. 8 in this volume).

On the other hand, if there were no abstract entities, it would certainly be mysterious that we proceeded as if they really played a role in scientific knowledge (mathematical and empirical) and philosophical elucidation. Therefore, even though defenders of their nonexistence (known as anti-realists about such objects or as nominalists) do not always think they have the burden of the proof—rather, on the contrary, they frequently argue that the burden is with the realists—at least they need to clarify the linguistic uses in which we seem to appeal to such entities, especially those given by assertions (or at least apparent assertions) that are commonly considered true, or to explain why those uses are not ontologically relevant. While defenders of abstract objects count on relatively simple ways to explain how true statements are evaluated in terms of those entities, detractors of their existence have either to justify the false character of such statements, or to provide an alternative (on certain occasions ad hoc) semantic analysis, or to state that such sentences are not really assertions but quasi-assertions.

Those who advocate for reinterpretations of mathematical statements that are compatible with their truth deserve separate consideration; for example, for a modal paraphrase of mathematics that aims at avoiding a commitment to mathematical entities while supporting their objectivity, their truth, see Putnam and Hellman (Putnam 1971; Hellman 1989).

Among the most important anti-realist or anti-platonist approaches, we can find (partially following Kalderon 2005; Eklund 2017) the following: (i) Those who consider that sentences about putative abstract objects have no content, and hence they (i.1) could be false, or (i.2) lack truth value. Bearing this in mind, proponents of this sort of view try to paraphrase sentences about putative abstract objects using sentences about supposed actual concrete objects and consider that sentences about abstract objects are mere instruments for establishing conclusions about concrete objects (Field 1980, 1989). (ii) Other anti-platonist approaches accept that sentences about putative abstract objects have content. Nevertheless, since the acceptance of this content cannot be equivalent to believing it, they urge some kind of fictionalism. According to one of these fictionalist approaches, (ii.1) the sentences accepted are quasi-assertions, and there is no need to evaluate them as true or false. Another way to go for those accepting that sentences about abstract objects have content is to claim that (ii.2) we do not have to believe these contents, whatever their truth value, because the function of these sentences is not that we believe them; rather, their function is quite another, for example, to provide *instrumental accounts* of true observational sentences (Field 1980) or to provide a fictional story to entertain us (see Yablo 1998). Some of these fictionalists also differentiate between the apparent and the real content of some sentences (see Yablo 2001, 2002, 2010 and 2014).

Also, besides defenders and detractors of abstract entities, there are those who adopt an agnostic position concerning such entities; a position that does not pronounce either in favor or against their existence: a position, in some sense, "neutral" between realism and anti-realism. Carnap's "Empiricism, Semantics and Ontology" (1950) can be interpreted in those terms. After Carnap's seminal article, several deflationist approaches were advocated, many of them vindicating the particular proposed interpretation to be a correct interpretation of Carnap's article (e.g., perhaps Putnam 1987, 1990; Hirsch 2002, 2011; Sider 2007, 2009; Thomasson 2015). Besides, there are deflationist proposals that do not vindicate Carnap's approach, among them, Azzouni's deflationist nominalism (Azzouni 2010) or Bueno's agnostic nominalism (Bueno 2008a, b and in this volume), which seems very similar to Carnap's view on abstract objects.

Contributions to this volume depict positions and debates that directly or indirectly involve taking one position or other about abstract objects of different kinds and categories. Their purpose is not to give a detailed account of the arguments deployed in the literature but to provide some samples of how positions for or against can be used in different areas of philosophy in relation to different matters.

The volume begins with an introduction by M. Plebani. The introduction is meant to present the main issues around the debate for those readers who are not familiar with it. This overview of the issue of abstract objects is structured around three questions: (1) What is an abstract object? (2) Why is the debate over the existence of abstract objects important? (3) How should we conduct the debate?

Other chapters in this volume, arranged in five parts, examine many interesting points connected with the three aforementioned questions in a variety of disciplines.

Part I, Enhanced Indispensability and Type Theories, encompasses a chapter by Zalta in which he argues for his relational typed object theory, and two chapters by Baron and by Martínez-Vidal and Rivas-de-Castro in which issues involving the latest and very sophisticated versions of the indispensability argument are addressed.

Indispensability arguments are presented in the Introduction as one of the available ways of approaching the debate about the existence of mathematical entities as abstract objects. In his chapter, S. Baron deals with one of the problems that have been posited to Baker-style versions of the enhanced indispensability argument. The latter aims to establish that mathematical entities exist because they are explanatorily indispensable for cases such as the cicadas. The argument applies inference to the best explanation to conclude that mathematical entities exist just as scientific realists infer the existence of non-observable entities. One of the difficulties confronted by Baker's proposal (not a crucial one according to Baron) is that the explanandum itself is formulated in mathematical terms; as a result, the truth of the explanandum is questioned, and the argument is contended not to obtain. Baron intends to overcome this situation by discussing a case of extra-mathematical explanation in which there is no mathematics in the explanandum. It is a case of bistable perception in which an ambiguous duck-rabbit image disambiguates one way or another in a switching process. He considers two models of mathematical explanation of the phenomenon; his intention is to conclude that mathematics is explanatorily indispensable because mathematical facts make a difference for physical facts even if mathematics is not needed to describe what is going on. Yet, the difficulty remains that a plausible nominalist account is also available. In order to overcome it, he suggests a strategy that might succeed to show that mathematics takes the explanatory load.

C. Martínez-Vidal and N. Rivas-de-Castro assess the latest developments around the enhanced indispensability argument. In particular, they discuss three positions (Baker 2017; Baron 2019; Knowles and Saatsi 2019) which agree that there is a substantive mathematical involvement in scientific explanation while differing on the way such involvement is to be understood and about the ontological commitments that derive from it. Baker and Knowles and Saatsi agree that the debate is at an impasse and intend their respective proposals to overcome it. Baker tries to do so by appealing to well-known explanatory virtues, while Knowles and Saatsi intend to surpass Baker's move by characterizing substantive mathematical involvement and its ontological commitments from the viewpoint of an independently well-established theory of explanation. Arguably, typifying it from an independent stance should prevent their proposal to be theoretically laden by their previous nominalist position. Nevertheless, Baron (2016, 2019) rejects the counterfactual theory can do the job and defines substantive mathematical involvement in terms of a modified version of the nomological account of explanation. The chapter appraises the dialectical move advanced by Knowles and Saatsi and the criteria for ontological commitment following from the three characterizations of substantial mathematical involvement to conclude that relying on independently established theories of explanation does not settle the issue because the choice of one or another theory of explanation is, after all, inspired by each author's preferred metaphysical view on mathematical entities. Moreover, the authors conclude that the explanatory version of the indispensability argument might not take the realist further than its more traditional version since it fails to establish that mathematical entities make a difference.

E. Zalta's contribution to this volume discusses and vindicates the way in which relations are understood in his relational type theory—which he calls typed object theory (TOT)—versus the way in which they are understood, respectively, in other relational type theories and in variants of the functional type theory. The purpose of the chapter is twofold; on the one hand, he discusses preexisting functional (Church 1940; Montague 1973) and relational theories (Orey 1959; Church 1974; Gallin 1975) to show that the main works on relational type theory took relations seriously only in the syntax; but semantically, they either preserved Frege's method of reducing relations to functions or took relations to be sets of n-tuples. Hence, one way or other, they all advocate a Fregean interpretation of predication that fails to capture what he takes to be the essential fact about it, namely, that “Fx” means that F characterizes x. Then he proceeds to argue that his TOT appropriately explains the phenomena that contemporary type theories (Muskens 2007; Liefke 2014; Liefke and Werning 2018; Williamson 2013) were designed to address while avoiding a number of problems they encounter.

Part II, Fictionalism or Realism in Philosophy of Mathematics, includes two chapters by O. Bueno and M. Leng. Both chapters reflect on what is the result of applying a naturalistic methodology to questions in the philosophy of mathematics.

O. Bueno argues for the unnecessary character of mathematical objects and mathematical truths. His contention is that an analysis of mathematical practice does not support any of the metaphysical thesis endorsed by the view he refers to as “the traditional view”: platonism understood as committing to the independent existence of abstract mathematical objects and to the necessary, if true, character of mathematical statements. He typifies the traditional view in front of other proposed ways of understanding platonism to continue to confront the way in which the view understands mathematical existence and mathematical necessity to the way in which these two features are to be understood in the light of mathematical practice. His conclusion is that mathematical practice does not support the way in which the traditional view conceives of the existence or necessity of mathematics. He proceeds to analyze the prospects of nominalism to conclude that mathematical practice does not suffice to establish that mathematical entities do not exist either; hence, he suggests agnosticism is the right way to go. Both nominalists and agnosticists grant the platonist that mathematical entities would be abstract if they existed, while both agree that mathematical practice seems to establish that mathematical entities are introduced by mathematical principles and those are not independent of us. In the final and more substantial part, the author argues for the contingent character of mathematics in the light of a detailed analysis of the notion of mathematical pluralism.

M. Leng opens her chapter with the Carnap-Quine controversy on the internal/external distinction. She follows Quine in his rejection of the distinction only to join Maddy (1997) to point out that it needs qualification. Quine contended that our pragmatic decisions turn into reasons to believe the assertions following from our chosen framework and into reasons to adopt metaphysical commitments, even to abstract entities; that would be so because abstract entities, mathematical entities among them, are indispensable to our best scientific theories. Maddy (1997) objected to Quine that scientists themselves take some parts of their theories literally while refraining to do so for other parts and that they do not commit to putative theoretical entities until they count on some sort of causal evidence for them. Then, Leng brings in Yablo’s (1998) and Azzouni’s (1998) criticisms of Quine’s proposal to rebut them. Leng claims that the challenges by these authors to Quine’s naturalistic approach to ontology are answered in a more meticulous naturalism like the one she offers in 2010. Nevertheless, Leng assumes that even the most meticulous naturalism is threatened by challenges such as those posed by Maddy (2011), when she claims that although there is nothing in science that supports robust realism, there is no fact of the matter to decide between Arealism and Thin Realism. Attending to this, Leng ends up accepting that the naturalist project can leave the answer to some questions about the existence of abstract objects as a matter of decision.

Part III, Fictionalism or Realism in Philosophy of Empirical Sciences, comprises several chapters that discuss the nature of scientific models (and in some of the chapters, the nature of scientific representation in general) as well as the need to appeal or not to abstract entities when accounting for/elucidating them.

J. Díez characterizes, in his chapter, the use of models—a characterization that applies to any other representational device—as a scientific practice consisting in the construction of "ensembles" and taking elements/parts/features of those ensembles as standing for other entities. This account, where a model (M) aims at its intended target system (T)—assumed the directionality from M to T—untangles two issues: (i) in virtue of what T is represented by M and (ii) in virtue of what that representation is successful. Finally, Díez argues that his account offers a clear advantage against other accounts of representational models, namely, it does not involve a commitment to the abstract character of scientific models. Yet, he does not reject that scientific models serve themselves of abstract objects, for instance, as their constituents.

X. de Donato-Rodríguez and J. L. Falguera argue that models are works of fiction though they are not fictional objects. They propose a view according to which a scientific model, as a work of fiction, should be understood as an intensional object which determines a fictional system; being both—scientific models and fictional systems determined by them—abstract artifacts. In this proposal, there are two other components to be considered: the material support(s) expressing a model and the target system (in case it exists) of the real concrete world. To vindicate their account, de Donato-Rodríguez and Falguera discuss two kinds of recent proposals about the nature of scientific models relating them with works of fiction: on the one hand, the approaches by Frigg, Toon, Levy, and Frigg and Nguyen, who accept that scientific models are or have to do with fictional objects when these are understood as nonexisting entities, and, on the other hand, an approach by Contessa, who proposes a dual ontological nature for scientific models and that the models (according to one of the two aspects of their dual nature) as imaginary objects are abstract objects. de Donato-Rodríguez and Falguera reject the first kind of proposals because they, in some sense, assume that scientific models are imaginings, though imaginings do not account for the intersubjective character of scientific models. They also reject Contessa's proposal because it advocates for a dual ontological nature, while it is more suitable to conceive a scientific model as just an abstract object, because it is an intensional entity of intersubjective character, which determines a fictional (or imaginary) system. Finally, de Donato-Rodríguez and Falguera try to show that their view presents advantages when it comes to account for the semantics of so-called external, internal, and intrafictional sentences related to scientific models.

S. Psillos contends that at least some scientific models are abstract entities. He helps himself with Carnot's machine case in order to make and illustrate his point. He acknowledges that models of the sort of Carnot's machine lack concrete instances and that descriptions associated with models are true descriptions of something; hence, since they describe something and they are not descriptions of concrete entities, they have to describe abstract objects. Psillos calls models "physical abstract entities" because they are abstract entities characterized by physical properties. Then he distinguishes three types of abstraction—Aristotelian, Newtonian, and Duhemian—and argues against fictionalist proposals (such as Frigg's) by trying to show that certain scientific models are the result of abstraction processes by means of which some properties are discovered, not just created (as it happens with fictional characters). Psillos needs to commit to what he takes to be a moderate scientific realism in order for this conclusion to obtain.

F. Salis, R. Frigg, and J. Nguyen elucidate how it is feasible to maintain that scientific models denote target systems while adopting an anti-realist conception about scientific models. The position has been charged with incoherence since anti-realism about scientific models pleads that models are like fictions and so nonexistent items and hence non-denoting entities. Their proposal considers models together with their descriptions understood as ways of prescribing imaginations (or model systems) about particular, possibly fictional (not concrete), systems. To address the issue, they appeal to the intentional character of thoughts that are directed at objects, even in those cases in which the intended objects do not exist and apply it to imaginations. Imaginations can be directed toward nonexistent objects just as they are sometimes directed toward existent (concrete) objects. Model M, as a combination of the description of model D and its content C, leads to an imagination that is directed at an imaginary object. Sometimes models denote target systems in the concrete world. However, contrary to what might initially be assumed, it would not be the imaginary models or model systems that would denote; rather, it is the description of the model and its content that denote. This approach is integrated into what elsewhere they have called the DEKI account of scientific representation (models *denote* their targets and *exemplify* certain properties, which are translated via a *key*, with properties to be *imputed* to their target).

Part IV, Fictionalism or Realism in Philosophy of Language, includes several chapters that discuss the problem of the existence of abstract objects in relation to topics of the philosophy of language arena.

M. García-Carpintero analyzes the problems that true claims of fictional co-identification emerge when trying to explain their truth conditions. The author revises several proposals about how to understand fictional names' contribution to the truth value of different kinds of utterances in which they occur. His preferences and arguments go with a proposal in line with Yablo's figuralism; according to it, the truth value of such utterances rests on pretenses of what García-Carpintero calls "textual uses of fictional discourses" (i.e., like those we directly encounter in novels). His main thesis in this chapter is that an adequate answer to the co-identification problem (identifying a character through different novels, for instance) must be consistent with a certain descriptivism he advocates.

S. Miguens argues for the concrete character of fictional characters. Her conclusion follows from a methodological premise: the analysis of the metaphysical status of fictional entities can be done in terms of some of our practices. In particular, she contends literature—hence fictional characters featuring in literary works—is essential to our moral practices, and that fact is incompatible with a commitment to their abstract character. Conceiving of the use of fiction and fictional characters in our moral practices to their role in arguments might be compatible with such abstract character; but she rejects that view. Argument needs concepts to be precise, but fiction goes beyond to help us re-carve our moral concepts. Our having, and our reflecting about having, one or other moral attitude or value involves our engaging with fictional characters that entertain them. Next, she introduces our contemporary view of the concrete-abstract distinction, pursues its historical origin, and claims that it is the result of the philosophical debate at the end of the nineteenth and beginning of the twentieth centuries. Meinong's view is especially relevant to her proposal, though she rejects Zalta's to subscribe to the sort of concrete neo-Meinongianism postulated by Parsons and argue that the only way of accounting for the role literature plays in our moral practices is by conceiving of fictional entities as concrete nonexistent objects.

In her contribution, F. Moltmann elaborates on her core-periphery distinction and in the defense of what she calls "the abstract object hypothesis". According to it, "natural language does not involve reference to abstract objects in its ontological core, but only in its ontological periphery." Her contention goes against the view that reference to abstract objects is "pervasive" in natural language. She claims that those authors who, like neo-Fregeans, appeal to natural language to argue for abstract objects are mistaken. The reason is that philosophers have traditionally relied on core uses of terms and other expressions purportedly referring to abstract objects to infer their existence, but core ontological uses of terms and expressions, that is to say, nonreflective uses of those terms in sentences (take, for instance, the use of "number" in "the number of homeless has increased a lot lately"), can be analyzed so as to do away with any commitment to them. On the contrary, peripheral ontological uses of terms such as "proposition" in "the proposition that 5 is greater than 2" do refer to abstract objects. But she contends that natural language ontology (be it core or periphery) is an ontology of appearances and one "that allows for considerable flexibility and expansion." Her favorite conception of abstract objects, as referents of peripheral ontological terms and as part of the ontology of natural language, is a pleonastic conception of abstract objects. Yet, she emphasizes that the semantic theory for those expressions needs to accommodate not only her preferred pleonastic view but be general enough to be compatible with other metaphysical views on abstract objects. Finally, in the appendix, she enumerates and rejects several objections that have been set to the abstract object hypothesis.

A. Voltolini introduces fictional creationism as the stance that contends that fictional entities are mind-dependent abstract entities; his aim is to argue for a restrained form of the view: moderate creationism. According to him, the moderate version of creationism provides the right answer to some recent criticisms to fictional creationism, namely, the ontological critique by Kroon or the metaphysical objections put forward by Deutsch and Brock. Moderate fictionalism contends that a make-believe practice is necessary but not sufficient for a *fictum* to acquire existence; in other words, mere imagination games do not allow for the creation of *ficta*. The existence of a *fictum* requires, in addition, a make-believe practice that involves a certain reflexive stance that identifies some set of properties with an individual postulated in a make-believe practice. As a result, "*ficta* can be seen to be stipulated entities qua the outcome of certain constitutive rules; they are not entities that are there before the stipulation." In addition, Voltolini defends a strong version of moderate creationism, according to which, the generation of a fictum has to do with constitutive rules (which are public) and not with supposed factors that cause it.

Finally, Part V, Fictionalism or Realism in Moral Philosophy and Philosophy of Arts, comprises three chapters addressing the abstract object matter in both those fields.

M. Balaguer advocates for "moral folkism," a view about moral properties which leads to a deflationary picture of many normative and metaethical questions (but not of applied ethical questions). It is deflationary because it contends that these questions are not settled by mind-independent facts about reality but by empirical facts about what ordinary folk happen to mean by the words that express those moral properties. Balaguer considers the case of "wrongness." He argues that sentences about wrong-like properties purport to be about abstract entities—so that if those properties exist at all, they are abstract objects; and he advocates that since there are multiple ways of conceiving of "wrong" (Kant-wrongness, Mill-wrongness, Moore-wrongness, etc.), if there are wrong-like properties, then there is a vast plurality of them (as a consequence of a plenitudinous platonism, the view that all the abstract objects that could exist actually do exist). Nevertheless, whether those properties exist is not so important to him, because his conclusions about moral facts obtain either way. The thing is that establishing that there is a myriad of ways of understanding the term "wrongness" does not settle the query about what counts as wrongness. According to him, this question is determined by our folk practices, by which of all those properties, if any, we have in mind when we use sentences with the term "wrong" in them. In his own words, "The metaphysical question is the semantic question." Yet, from the fact that what we mean by "wrong" depends on our practices, it does not follow that the facts that decide about applied moral facts are semantic facts. Those have to do with whether there is a property that those facts instantiate. From the fact that there is a certain wrongness property, it does not follow that this property is instantiated. Finally, Balaguer argues that his moral folkism is a sort of weak realism; nevertheless, if he is wrong and some of his initial assumptions are false (for instance, if we do not use "wrong" as property attributing), then some sort of cognitivism or expressivism might be true.

E. Caldarola embraces a version of hermeneutical fictionalism, a meta-ontological way of understanding works of art that is a version of Yablo's semantic descriptivism. Yablo sustains that there are certain texts/utterances for which we need to distinguish their real and literal content because only their real content is relevant to determine their ontological commitments. For instance, when we say "the number of homeless increases," the real content of the utterance is not about the existence of numbers but about homeless people. Analogously, we should distinguish between the real and the literal content of our performances of songs and symphonies and works of performance art and installation art: the real content of those performances (for instance, listening to a song) has nothing to do with the existence of works of art understood as abstract objects. Thus, our saying "I must have listened to Ani Franco singing Swim" has nothing to do with talking about an abstract object. Caldarola follows Yablo to argue that there is a functional mechanism that links the real content of a statement with its literal content and follows Walton to propose that there are principles of generation in games of make-believe that link what is true in the real world to what is true in the fictional world. She starts introducing the view above to continue presenting four other meta-ontological views of works of art and the problems that have been posed to each of them in the literature. She argues that Dodd's folk-theoretic modesty and David Davies' rational accountability are to be preferred—even if they are not without problems—to Thomasson's and Kania's descriptivist views. Finally, she discusses the advantages of her hermeneutic fictionalism about works of art; in particular, she claims that it is neither revisionary about our commonsense view of works of art nor an error theory about works of art.

C. Jay presents an argument for moral fictionalism that the moral realist can also accept. He assumes that moral facts do not depend on moral objects (in the sense in which ontological mathematical realism does) and that our beliefs are morally relevant because they are action guiding. To formulate his argument, he distinguishes belief from non-doxastic commitment, objective versus subjective warrant, hyper-enkratic beliefs, and sub-enkratic beliefs, acting on the basis of something we wish to be true versus acting on the basis of something that we desire and does not obtain but we commit to. Armed with this conceptual tools, he concentrates in those cases in which a person fails to form an intention to act on their belief (despite their fully believing) because they lack subjective warrant for that belief. In those cases, the realist proposes that we should try to change that condition into a situation in which we do have a subjective epistemic moral warrant. On the contrary, the fictionalist contains that our moral attitudes are non-doxastic attitudes toward morality, psychological states in which we accept (not believe) certain moral claims. Jay's point is that in those cases in which we lack subjective epistemic warrant—considering those cases in which our moral commitments are good—non-doxastic moral commitments can appropriately guide our action. That would be so because if, as the moral realist contends, moral beliefs guide our action, in those cases in which we lack epistemically (and subjectively) warranted moral beliefs, we will fail to act well. But if we allow acting based on non-doxastic morally good attitudes, then we will act well. Finally, he distinguishes the position he advocates, evaluative fictionalism, from revolutionary fictionalism. Only the former is compatible with moral realism because it does not commit to the thesis that "we ought to make our moral commitments non-doxastic."

References

Azzouni, J. (1998). On "On what there is". *Pacific Philosophical Quarterly, 79*(1), 1–18.

Azzouni, J. (2010). *Talking about nothing: Numbers, hallucinations, and fictions*. Oxford: Oxford University Press.

Baker, A. (2017). Mathematics and explanatory generality. *Philosophia Mathematica 25*(2), 194–209.

Balaguer, M. (1998). *Platonism and anti-platonism in mathematics*. New York: Oxford University Press.

Baron, S. (2016). Explaining mathematical explanation. *The Philosophical Quarterly, 66*(264), 458–480. https://doi.org/10.1093/pq/pqv123.

Baron, S. (2019). Mathematical explanation by law. *British Journal for the Philosophy of Science, 70*(3), 683–717. https://doi.org/10.1093/bjps/axx062.

Bueno, O. (2008a). Nominalism and mathematical intuition. *Protosociology, 25*, 89–107.

Bueno, O. (2008b). Truth and proof. *Manuscrito: revista internacional de filosofía, 31*(1), 419–440.

Carnap, R. (1950). Empiricism, semantics, and ontology. *Revue Internationale de Philosophie, 4*(11), 20–40.

Church, A. (1940). A formulation of the simple theory of types. *The Journal of Symbolic Logic, 5*(2), 56–68.

Church, A. (1974). Russellian simple type theory. *Proceedings and Addresses of the American Philosophical Association, 47*, 21–33.

Colivan, M. (2001). *The indispensability of mathematics*. New York: Oxford University Press.

Eklund, M. (2017). Fictionalism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2017 Edition). https://plato.stanford.edu/archives/win2017/entries/fictionalism/

Field, H. (1980). *Science without numbers.* Princeton: Princeton University Press.

Field, H. (1989). *Realism, mathematics and modality*. Oxford: Blackwell.

Frege, G. (1884). *Die Grundlagen der Arithmetik*. Breslau: Wilhelm Koebner. (English translation: *The Foundations of Arithmetic*. Oxford: Blackwell, 1959).

Gallin, D. (1975). *Intensional and higher-order modal logic: With applications to Montague semantics*. Amsterdam: North-Holland.

Hale, B. (1988). *Abstract objects*. Oxford: Blackwell.

Hellman, G. (1989). *Mathematics without numbers.* Oxford: Clarendon Press.

Hirsch, E. (2002). Quantifier variance and realism. *Philosophical Issues, 12*, 51–73.

Hirsch, E. (2011). *Quantifier variance and realism.* New York: Oxford University Press.

Kalderon, M. E. (2005). *Moral fictionalism*. Oxford: Clarendon Press.

Knowles, R., & Saatsi, J. (2019). Mathematics and explanatory generality: Nothing but cognitive salience. *Erkenntnis*, 1–19. https://doi.org/10.1007/s10670-019-00146-x.

Leng, M. (2002). What's wrong with indispensability? *Synthese, 131*(3), 395–417. https://doi.org/10.1023/A:1016141509719

Lewis, D. (1986). *On the plurality of worlds*. Oxford: Basil Blackwell.

Liefke, K. (2014). *A single-type semantics for natural language*. Enschede: Ipskamp Drukkers.

Liefke, K., & Werning, M. (2018). Evidence for single-type semantics – an alternative to *e/t*-based dual-type semantics. *Journal of Semantics, 35*(4), 639–685.

Maddy, P. (1997). *Naturalism in mathematics*. Oxford: Oxford University Press.

Maddy, P. (2011). *Defending the axioms*. Oxford: Oxford University Press.

Moltmann, F. (2013). *Abstract objects and the semantics of natural language.* Oxford: Oxford University Press.

Montague, R. (1973). The proper treatment of quantification in ordinary english. In J. Hintikka, J. Moravcsik, & P. Suppes (Eds.), *Approaches to Natural Language: Proceedings of the 1970 Stanford Workshop on Grammar and Semantics* (pp. 221–242). Dordrecht: D. Reidel. Reprinted in R. Thomason (Ed.). (1974). *Formal philosophy: Selected papers of richard Montague* (pp. 247–270). New Haven: Yale University Press.

Muskens, R. (2007). Intensional models for the theory of types. *The Journal of Symbolic Logic, 72*(1), 98–118.
Orey, S. (1959). Model theory for the higher order predicate calculus. *Transactions of the American Mathematical Society, 92*, 72–84.
Putnam, H. (1971). *Philosophy of logic*. New York: Harper and Row.
Putnam, H. (1987). *The many faces of realism*. LaSalle: Open Court.
Putnam, H. (1990). *Realism with a human face*. Cambridge, MA: Harvard University Press.
Quine, W. V. (1948/1953). On what there is. *The Review of Metaphysics, 2*(5), 21–38. And in: *From a logical point of view* (p. 1–19). Cambridge, MA: Harvard University Press.
Quine, W. V. O. (1969). *Ontological relativity and other essays*. New York: Columbia University Press.
Rosen, G. (2008). Abstract objects. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2018 Edition). https://plato.stanford.edu/archives/win2018/entries/abstract-objects/
Sider, T. (2007). Neo-fregeanism and quantifier variance. *Aristotelian Society, supplementary volume, 81*, 201–32.
Sider, T. (2009). Ontological realism. In D. Chalmers, D. Manley, & R. Wasserman (Eds.), *Metametaphysics* (pp. 384–423). Oxford: Oxford University Press.
Thomasson, A. (1999). *Fiction and metaphysics*. Cambridge: Cambridge University Press.
Thomasson, A. (2015). *Ontology made easy*. New York: Oxford University Press.
Williamson, T. (2013). *Modal logic as metaphysics*. Oxford: Oxford University Press.
Wright, C. (1983). *Frege's conception of numbers as objects*. Aberdeen: Aberdeen University Press.
Yablo, S. (1998). Does ontology rest on a mistake? In *Aristotelian Society Supplementary Volume, 72*(1), 229–283.
Yablo, S. (2001). Go figure: A path through fictionalism. *Midwest Studies in Philosophy, 25*(1), 72–102. https://doi.org/10.1111/1475-4975.00040
Yablo, S. (2002). Abstract objects: A case study. *Philosophical Issues, 12*(1), 220–240. https://doi.org/10.1111/j.1758-2237.2002.tb00068.x
Yablo, S. (2010). *Things: Papers on objects, events, and properties*. New York: Oxford University Press.
Yablo, S. (2014). *Aboutness*. Princeton: Princeton University Press.
Zalta, E. N. (1983). *Abstract objects: An introduction to axiomatic metaphysics*. Dordrecht: D. Reidel.
Zalta, E. N. (1989). *Intensional logic and the metaphysics of intentionality*. Cambridge, MA: Bradford Books/The MIT Press.

Santiago de Compostela, Spain

José L. Falguera
Concha Martínez-Vidal

Contents

Contributors

Mark Balaguer Department of Philosophy, California State University, Los Angeles, CA, USA

Sam Baron University of Western Australia, Crawley, WA, Australia

Otávio Bueno Department of Philosophy, University of Miami, Coral Gables, FL, USA

Elisa Caldarola University of Padova, Padua, Italy

Xavier de Donato-Rodríguez University of Santiago de Compostela, Santiago, Spain

José A. Díez LOGOS Research Group, Barcelona Institute of Analytic Philosophy, University of Barcelona, Barcelona, Spain

José L. Falguera University of Santiago de Compostela, Santiago, Spain

Roman Frigg London School of Economics and Political Science, London, UK

Manuel García-Carpintero Departament de Logos-BIAP, Department of Philosophy, University of Barcelona, Barcelona, Spain

Christopher Jay Department of Philosophy, University of York, York, UK

Mary Leng Department of Philosophy, University of York, York, UK

Concha Martínez-Vidal Department of Philosophy and Anthropology, University of Santiago de Compostela, Santiago de Compostela, Spain

Sofía Miguens University of Porto, Porto, Portugal

Friederike Moltmann CNRS, Paris, France

James Nguyen Department of Philosophy, University College London, London, UK
Institute of Philosophy, School of Advanced Study, University of London, London, UK
Centre for Philosophy of Natural and Social Science, London School of Economics and Political Science, London, UK

Matteo Plebani University of Torin, Turin, Italy

Stathis Psillos Department of History and Philosophy of Science, University of Athens, Athens, Greece

Navia Rivas-de-Castro University of Santiago de Compostela, Santiago, Spain

Fiora Salis Department of Philosophy, University of York, London, UK

Alberto Voltolini Department of Philosophy and Education Sciences, University of Turin, Turin, Italy

Edward N. Zalta Center for the Study of Language and Information, Stanford University, Stanford, CA, USA

Chapter 1
Recent Debates over the Existence of Abstract Objects: An Overview

Matteo Plebani

Abstract This volume is dedicated to recent debates over the existence of abstract objects. Three relevant questions for such debates are: (1) what is an abstract object? (2) Why is the debate over the existence of abstract objects important? (3) How should we conduct the debate? (See Burgess and Rosen 1997, p. 12 for a list of similar questions).

This volume is dedicated to recent debates over the existence of abstract objects. Three relevant questions for such debates are: (1) what is an abstract object? (2) Why is the debate over the existence of abstract objects important? (3) How should we conduct the debate? (See Burgess and Rosen 1997, p. 12 for a list of similar questions).

In this overview, we will survey traditional answers to those questions (readers already familiar with the debate can skim or skip this chapter).

1.1 What Is an Abstract Object?

One way to answer this question is by simply giving a list of examples of abstract objects. Prototypical examples of abstract objects include mathematical objects (numbers, functions, sets . . .), universals (the property of being red, the property of being round . . .), propositions (conceived as the contents of sentences and the objects of belief), abstract types as opposed to concrete tokens (Marx's *Das Kapital* as opposed to Lenin's copy of *Das Kapital*, see Wetzel 2009).[1] Other candidates to the title of abstract object are fictional characters like Sherlock Holmes (Van

[1]It is standard to assume that no abstract object is concrete. However, some authors (Williamson 2013) hold that some objects are neither abstract nor concrete.

M. Plebani (✉)
University of Turin, Turin, Italy
e-mail: matteo.plebani@unito.it

J. L. Falguera, C. Martínez-Vidal (eds.), *Abstract Objects*, Synthese Library 422,
https://doi.org/10.1007/978-3-030-38242-1_1

Inwagen 1977; Thomasson 1999), scientific models (Contessa 2010), moral values (Mackie 1977, see also the list of candidates for abstractness in Liggins 2010), artworks (Mag Uidhir 2013) and perhaps institutional entities like universities and marriages (Burgess and Rosen 1997, p.15). The category of abstract objects is supposed to stand in contrast to the category of concrete objects; paradigmatic examples of concrete objects are human beings, animals, houses, stars and electrons.

When one tries to define the notion of abstract objects by giving a list of examples of abstract and non-abstract objects, the lists provided are not supposed to be exhaustive: the examples are offered to suggest that an abstract object is one like those in the list and one that differs in some important respects from the examples of concrete objects offered. However, isolating the features that distinguish abstract and concrete objects is a difficult task.

The popular way to approach the abstract/concrete distinction is to focus on some features that concrete objects possess but abstract objects lack (Lewis 1986, §1.7, calls this definition of abstractness "the way of negation". See also Rosen 2017). According to the standard definition, an object is abstract if and only if it has no spatial location and has no causal relations (abstract objects do not make things happen and there is no way to act upon them).

The definition has initial plausibility, given that it does not seem to make sense to say that numbers exist here but not there and now but not then; and the opinion that numbers are a-causal seems to be confirmed by the fact that we cannot perceive them in any way and mathematicians do not study abstract objects by making them interact with other physical objects which possess causal powers (See Linnebo 2018, section 2.5 for a similar suggestion).

Despite its initial plausibility, there seems to be some counterexamples to the standard definition of abstract object (see Rosen 2017). As a counterexample to the idea that all abstract objects lack a spatial location, some argue that impure sets (sets whose members, or members of their members, or members of the members of ... their members are concrete objects) share the location of the concrete elements they contain: on this account, {Obama} is located where Obama is (Lewis 1986). As counterexamples to the a-causality of abstract objects one could cite the case of what could be called *abstract artifacts*. The commonsensical view about an abstract object like the game of chess is that it did not exist until it was invented (Rosen 2017). Similarly, the natural view on fictional characters like Sherlock Holmes is that they have been created by the writers of certain novels (Thomasson 1999 and Rosen 2017). In this sense fictional characters and games like chess are artifacts, objects that exist thanks to the activity of some human beings.

Moreover, is not even clear that abstract objects like books (conceived as types) do not have causal powers: *Das Kapital* arguably had an impact on twentieth century history. More generally, against the supposed a-causality of abstract objects, one might hold that the *relata* of causal relations are events and that abstract objects might be involved in some events that caused other events (Rosen 2017). One might hold that Pythagoras' theorem is part of the event in which I correctly remember Pythagoras' theorem and give the correct answer to a certain math question, thereby passing a math test.

In light of the difficulty of characterizing the notion of abstract object, some authors prefer to focus on particular examples of abstract objects. Field (1980), one of the most important recent defenses of a nominalist position, makes clear that its focus is on mathematical objects. Dorr (2008), Melia (2008) and Szabó (2003) choose to discuss mathematical objects and universals, in order not "to be drawn into a pointless debate about how to define that technical term [abstract]" (Dorr 2008, p. 34).

The choice to focus just on some particular examples of abstract objects instead of discussing in full generality the question of the existence of abstract objects seems legitimate. After all, the issue whether numbers and sets exist is interesting in itself and so is the question whether universals exist.

Moreover, there is an additional reason for those who want to deny the existence of some abstract objects not to formulate their position in too general terms. The problem with simply denying that there are any abstract objects is that apparent reference to abstract objects is ubiquitous in ordinary language, which makes the enterprise of denying the existence of every kind of abstract object we apparently refer to a formidable task. Some philosophers use such an apparent reference to abstract objects to make jokes about nominalism, the philosophical view that there are no abstract objects:

> if intellectual debts are abstract entities—and what else could they be? — then the fact that I have so many of them ought to serve by itself as an altogether convincing refutation of nominalism. (Rosen, preface to Burgess and Rosen 1997).

One could even detect an apparent reference to abstract objects in the standard definition of Nominalism as "the doctrine that there are no abstract entities" (Field 1980, 1): aren't doctrines abstract objects? (Cfr. Plebani MS).

Of course, no one sympathetic to the spirit of nominalism would be impressed by the aforementioned refutations of nominalism, which are better interpreted as jokes rather than serious arguments. But the fact remains that the list of potential candidates to the title of abstract object is very large. It makes little sense for an author interested in arguing against the existence of sets and numbers to discuss the question whether a nominalist account of doctrines or intellectual debts is available. This might justify the choice to focus on particular examples of abstract objects when discussing the question whether abstract objects exist.

Yet there is one reason to try to find a general account of abstract objects. The reason why many philosophers reject abstract objects is that they believe that our knowledge of abstract objects seems to be problematic in a way in which our knowledge of concrete objects is not (Burgess and Rosen 1997, 11–12). One naturally wonders which distinctive feature of abstract objects makes our knowledge of them so problematic. The following section explores this issue.

1.2 Why Debating over the Existence of Abstract Objects?

One reason to discuss the existence of abstract objects is that on the one hand, as we saw, apparent reference to abstract objects is ubiquitous in natural language;

on the other hand, some philosophers deny the existence of abstract objects. It is worth discussing which one of the following project is more promising: (i) try to develop an account of our ordinary and scientific discourse that does not posit abstract objects or (ii) show that the arguments presented against the assumption that there are abstract objects are not satisfactory?

In this section, I will focus on the prospects for project (ii). I will start by reviewing the reasons that philosophers have offered to believe that there are no abstract objects. In their seminal paper, Goodman and Quine (1947) claim that the rejection of abstract objects "is based on a philosophical intuition that cannot be justified by appeal to anything more fundamental" (Goodman and Quine 1947, p. 105). It might well be true that philosophers of a certain temperament tend to have antipathy towards abstract entities.[2] However, intuitions vary among philosophers. What Goodman and Quine call a primitive intuition other philosophers might well call a prejudice against abstract objects.

If one looks for a motivation for nominalism that goes beyond a mere appeal to intuition, one often finds arguments based on epistemological considerations. The term commonly used to refer to such considerations is "epistemological challenge against platonism", where platonism is the view that there are abstract objects (see Liggins 2010). At the root of such arguments is the idea that the a-causal character of abstract objects makes it very hard for those who believe that abstract objects exist to explain the correlation between our beliefs about those objects and truths about them. The strongest formulation of the challenge is considered the one offered by Hartry Field (1989, Introduction), which is inspired by Benacerraf (1973). Field notes that platonists admit a phenomenon that he calls "mathematical reliability": most of the mathematical beliefs held by the mathematicians are correct and that is not by accident. Arguably, as Liggins (2010, 70) notes, even most of the mathematical beliefs of many non-mathematicians are non-accidentally true. Mathematical reliability is "so striking to demand explanation" (Field 1989, 26), but according to Field "the claims the platonist makes about mathematical entities appear to rule out any reasonable strategy for explaining the systematic correlation in question" (1989, 231).

Field acknowledges that a first step towards explaining mathematical reliability is recognizing that in many cases mathematicians believe a certain theorem because they have derived it from the standard axioms. Assuming that we have an account of the mathematicians' logical competence, i.e. their ability to recognize when something logically follows from the axioms, this reduces the problem of explaining mathematical reliability to the problem of explaining the ability of the mathematicians to believe true axioms. Field challenges the platonist to explain this aspect of the phenomenon of mathematical reliability. Field also grants that it is possible to explain the mathematicians' ability to choose consistent sets of axioms; but he contends that this does not completely solve the problem of accounting

[2]It is telling that even later in his life, when he changed his mind on the issue of nominalism and came to accept abstract objects, Quine still referred to abstract objects as *entia non grata* (see Quine 1960, chapter 50).

for mathematical reliability because “there is a gap between the consistency of an axiomatic theory and its truth” (1989, 233).

A causal explanation of mathematical reliability seems to be ruled out by the a-causal nature of abstract objects. Platonists who accept the standard definition of abstract objects cannot claim that abstracts objects cause our beliefs about them, nor that beliefs about mathematical objects and truths about such objects have a common cause.

A non-causal explanation of mathematical reliability cannot be ruled out a-priori, Field admits, but he expresses some skepticism about the prospects of finding such a non-causal account of reliability. According to Field, part of the difficulty for such a non-causal explanation of mathematical reliability lies in the fact that mathematical objects are usually characterized by the platonists not only as a-causal, but also as mind-independent and language-independent. Field does not present his argument as a definite refutation of platonism, but rather as a challenge for the platonist, while at the same time offering some reasons to think that it is hard for the platonist to meet the challenge (see Liggins 2006, 2010). It should be noted that the challenge does not rely on any theory of knowledge. Field does not make any assumptions about the necessary or sufficient conditions for knowledge. His challenge is simply based on the idea that “we should view with suspicion any claim to know facts about a certain domain of objects if we believe it impossible in principle to explain the reliability of our beliefs about that domain” (1989, p. 233).

It should be noted that Fields’ epistemological challenge is a challenge to traditional forms of mathematical platonism: in other words, it is a challenge to the assumption that there are mind-independent and language-independent abstract objects. It is not so clear whether the challenge is equally forceful against those abstract objects that in the previous section we called *abstract artifacts*, which can be described as mind-dependent (at least in some sense).[3]

An important to point to make is that there are two ways in which philosophers have tried to meet the epistemological challenge. One response can be roughly characterized as based on the idea that the way we know about abstract mathematical objects is different from the way we know about concrete objects. Kit Fine endorses this view in the course of defending an account of our knowledge of mathematical objects called *procedural postulationism*, according to which abstract objects might be introduced in the domain of discourse by an act of postulation, an order to expand the domain by including objects of a certain kind, as long as the order is consistent. As he explains:

> Each kind of object has its own way of being known. It is a peculiarity of perceptible objects that we may get to know of them through perception; it is a peculiarity of the theoretical entities of science that their existence is to be justified by way of inference to the best explanation; and it is a peculiarity of mathematical and other abstract objects that their existence is to be justified by way of postulation. [. . .] If the present approach has any

[3] On the possibility of generalizing Field’s challenge beyond the case of abstract mathematical objects, see Liggins (2010) and Enoch (2010). See Rosen (1994) for problems with the distinction between mind-dependent and mind-independent, objects.

> value, it lies in its making clear the distinctive way in which we may acquire our knowledge of mathematical objects, one that is not reducible to other, more familiar methods and is in keeping with the peculiarly a priori character of mathematical thought. (Fine 2005, p. 108)

One way to spell out the difference between the epistemology of abstract mathematical objects and concrete objects is to say that in the case of mathematical theories, i.e. theories about mathematical objects, there is no gap between consistency and truth: *every consistent mathematical theory is true of some mathematical objects* (call the principle in italics the consistency-truth link). The consistency-truth link goes back at least to Hilbert and is present in some form also in Fine's account: the only constraint on the use of an act of postulation to introduce some mathematical objects is that the act of postulation be a consistent order.

The consistency-truth link is also present in the so-called *plenitudinous platonism* (Balaguer 1995, 1998), according to which every possible mathematical object exists. According to plenitudous platonism, it is sufficient for a mathematical object to exist that the assumption of its existence is part of a consistent set of axioms. The same is not true of physical objects and physical theories: consistency is not the criterion for truth and existence in physics, as Balaguer himself acknowledges. Similarly, in Fine's account only abstract objects might be postulated into existence, not concrete ones.

Zalta's theory of abstract objects (see Zalta 2016), arguably the most general account of abstract objects available, makes an important distinction between abstract and concrete objects, which has repercussions about the way we can know about the two different kinds of objects. Zalta distinguishes between two modes in which a property can be ascribed to an object: the notation Fa indicates that the object a exemplifies the property F, whereas the notation aF is used to indicate that the object a encodes the property F, in the sense that a is characterized as having property F. Flesh and blood detectives exemplify the property of being a detective, whereas Sherlock Holmes encodes the property of being a detective, in the sense of being characterized as possessing such property. In Zalta's account, concrete, i.e. spatio-temporally located, objects cannot encode properties; only abstract objects encode properties. On the other hand, a powerful comprehension principle guarantees that, for any group of properties, some abstract object encodes exactly those properties. Moreover, according to Zalta's theory, abstract and concrete objects have different identity criteria: concrete object x is identical to concrete object y if and only if x and y necessarily *exemplify* the same properties, whereas abstract object x is identical to abstract object y if and only if x and y necessarily *encode* the same properties. The reliability of our beliefs about mathematical objects, on this account, is explained by the fact we can recognize the properties that abstract objects encode, because they encode precisely those properties that are used to characterize them[4]:

[4]Compare Zalta's view that an abstract object encodes exactly those properties that are used in its characterization with Yablo's idea (Yablo 2010, Introduction) that mathematical objects are *preconceived* objects, objects that "Either ... *should* have feature F, given their job description, or ... *don't* have feature F" (Yablo 2010, 7).

> All one has to do to become so acquainted *de re* with an abstract object is to understand its descriptive, defining condition, for the properties that an abstract object encodes are precisely those expressed by their defining conditions (Linsky and Zalta 1995, p. 547).

In the case of the mathematical objects postulated in the mathematical theory T, such objects encode a certain property if and only if according to T they possess such a property, so that mathematical reliability can be explained by our capacity to identify the consequence of certain mathematical theories.

We have just seen that according to one reply to the epistemological challenge the explanation of the reliability of our mathematical beliefs has to be different from the explanation of the reliability of our perceptual beliefs or of our beliefs about theoretical entities of physics. A different reply to the epistemological challenge is possible, which is based on the idea that:

> we have the same broadly inductive reason for believing in numbers as we have for believing in electrons: certain theories that entail that there are numbers are better, qua explanations of our evidence, than any theories that do not. (Dorr 2010, p. 133)

> [I]f our belief in electrons and neutrinos is justified by something like inference to the best explanation, isn't our belief in numbers and functions and other mathematical entities equally justified by the same methodology? After all, the theories that we use in explaining various facts about the physical world not only involve a commitment to electrons and neutrinos, they involve a commitment to numbers and functions and the like. (Field 1989, p. 16)

The consideration that our best scientific theories are heavily mathematized is at the core of the so-called indispensability argument (see next section) for the existence of abstract mathematical objects. This suggests one way in which numbers and electrons might be assimilated: both are posited by our best scientific explanations, including explanations of empirical phenomena (Baker 2005, see also next session).[5] Some go as far as to claim that the central role played by abstract mathematical objects in our empirical theories makes it hard to "clarify the distinctive way in which ordinary material bodies are causally active" (Burgess and Rosen (1997), p. 23) and maintain that "abstracta [. . .] are not causally active in that way" (Burgess and Rosen (1997), p. 23). This strategy might be presented as an attempt to call into question the assumption that abstract mathematical objects are a-causal, an assumption used by Field to rule out the possibility of a causal explanation of mathematical reliability. In connection with the issue of the a-causality of abstract objects it is worth noticing that according to one prominent account of universals (Armstrong 1978) universals, traditionally classified as abstract objects, are considered causally active.

More generally, as Burgess and Rosen (1997, 31) note, the difficulty of defining the notion of abstract object might be used to motivate an account that posits objects traditionally classified as abstract (like mathematical objects or universals), but

[5] "In recent times, many philosophers have been attracted to an 'assimilationist' model of mathematical knowledge; they have supposed that we know of mathematical objects in something like the way we know of other objects" Fine (2005, p. 108–9).

avoids the claim that such objects belong to a category of objects which are radically different from ordinary concrete objects. In connection with the idea of blurring the abstract/concrete distinction, Alan Baker has raised the question whether

> Is it possible to clarify the distinctive way in which ordinary material bodies can play an explanatory role, and if so can it indeed be said that abstracta which are mentioned in the context of scientific explanations are not explanatory in that way (Baker 2009, p.632)[6]

It is worth noticing the connection between these two different replies to the epistemological challenge and the issue of whether it is possible to provide a satisfactory account of the abstract/concrete distinction. As we saw, one reply to the epistemological challenge is based on the idea that different kinds of objects are associated with different ways of knowing about them. This reply seems to presuppose that abstract and concrete objects are indeed different kinds of objects and that it is possible to trace the abstract/concrete divide in a satisfactory way.

The reply to the epistemological challenge based on the role of mathematics in our empirical theories, on the other hand, sits well with skepticism about the possibility of tracing the abstract/concrete divide in a satisfactory way. The two replies seem also to yield different criteria for the acceptability of a mathematical theory. According to the reply based on the consistency-truth link, the only criterion for the acceptability of a mathematical theory is its consistency, whereas according to the reply based on scientific confirmation the applicability of a theory to the study of the concrete world plays also an important role, so much so that Quine at a certain point considered the parts of set theory which have no connection to applied mathematics mere “mathematical recreation” (Quine 1986, p. 400) and suggested not to accept the existential implications of those parts of set theory.

We have found a connection between the issue of defining the notion of abstract object and the question of how to reply to the epistemological challenge against platonism: one reply to the epistemological challenge is based on a clear distinction between the way we know about abstract objects and the way we know about concrete ones and seems to presuppose a sharp distinction between abstract and concrete objects; while another reply argues that we have the same grounds for believing in the existence of prototypical abstract objects like numbers as we have for believing in the existence of prototypical concrete objects like electrons. This second reply is also in harmony with the idea that there is no sharp boundary between abstract and concrete objects, a view that receives some support from the difficulty of characterizing the notion of abstract object discussed in the previous section.

[6]See also Morrison (2007, 552, quoted by Knowles and Liggins (2015), 3406) on the property of spin: according to Morrison, spin is “perhaps best viewed as a curious hybrid of the mathematical and the physical” (Morrison (2007: 552) quoted by Knowles and Liggins (2015: 3406)).

1.3 How Should We Conduct the Debate?

One might be tempted to think that philosophical debates like the one over the existence of abstract mathematical objects are unsolvable, in the sense that there is no evidence capable of convincing someone to change opinion about such questions. Quine was of a different advice:

> Existence statements in this philosophical vein do admit of evidence, in the sense that we can have reasons, and essentially scientific reasons, for including numbers [in our ontology] . . . Numbers and classes are favoured by the power and facility which they contribute to theoretical physics and other systematic discourse about nature. (Quine 1969, pp. 97-8)

A good part of the recent debate over the existence of abstract mathematical objects can be presented as an attempt to figure out whether the passage from Quine just quoted is right or not. Do we have "essentially scientific reasons" (Quine 1969: 97–8) to believe that numbers exist? Those who answer *yes* usually subscribe to what is called "the Quine-Putnam indispensability argument".[7] At the core of the indispensability argument lies the remark that contemporary scientific theories are highly mathematized: such theories make apparent reference not only to concrete objects like neutrinos and magnetic fields, but also to abstract objects like functions and numbers. According to the argument, given that our best scientific theories entail the existence of abstract mathematical objects and we accept those theories, we should believe that abstract objects exist. The indispensability argument (henceforth IA) in its simplest form can be presented like this (see Liggins 2016: 532 and Plebani 2018: 255):

> Mathematics is indispensable to science: our best scientific theories entail the existence of abstract mathematical objects.
> If mathematics is indispensable to science, then there are mathematical objects.
> Therefore: there are mathematical objects.

If one does not want to call into question the validity of *modus ponens*, then to reject the argument one must reject one of its two premises. Field's defense of nominalism (Field 1980, 1989) is an attempt to reject premise (1) and provide a nominalistic versions of our best scientific theories that do not contain any reference to or quantification over abstract mathematical objects like numbers or sets. Field's project is to prove that there can be a *Science without Numbers* and so that mathematics is not indispensable to our best scientific theories. While Field's nominalistic version of Newtonian gravitational theory is considered an important accomplishment, there seems to be no clear way to extend the approach adopted in Field (1980) to provide a nominalistic reconstruction of contemporary physics (Malament (1982)). In light of this, the received wisdom is that at the moment Field's project has not been completed. In light of the difficulty posed by the task of

[7]Liggins (2008) argues that neither Quine nor Putnam actually endorsed the Quine-Putnam indispensability argument. See also Putnam (2012).

reformulating our scientific theories in a nominalistically kosher way, the reply to IA based on a denial of premise 1 goes under the name of "hard road nominalism" (Colyvan 2010; Liggins 2016).

A different reply to IA rejects its second premise. The reply grants that we need to formulate our theories mathematically; but it contends that this does not imply that mathematical objects exist. There might be various ways in which "numbers and classes . . . contribute to theoretical physics and other systematic discourse about nature" (Quine 1969, 97–8). Perhaps presupposing that numbers exist helps us to represent certain circumstances; this does not mean that numbers need to be there for those circumstances to obtain, or that we need to assume the existence of numbers to explain why those circumstances do obtain.[8] One way to reject premise two of IA is to argue that the role of mathematics in our scientific theories is merely that of a "representational aid" (Yablo 2005):

> mathematics appears in our empirical theories as a mere descriptive aid: by speaking in terms of the real number line, . . . or some other mathematical structure, we simply make it easier to say what we want to say about the physical world. (Balaguer 1998, p. 137)

> [T]he kind of theoretical utility that mathematics brings is not of the right kind [to justify belief in abstract mathematical objects] [. . .] Using numbers to index quantities may enable us to say much more complicated things about relationships between the various quantities, but it is nothing more than a labelling device. (Melia 2008, p. 117. See also Melia 2000)

This style of response to IA is sometimes also called "easy road nominalism" (Colyvan 2010) because it does not require the nominalist to reformulate our scientific theories. Some friends of IA have tried to argue against this kind of reply by pointing out that in some cases mathematics does not merely play a representational role, but also an explanatory one (Colyvan 2002; Baker 2005; Baker and Colyvan 2011). What goes under the name of "enhanced indispensability argument" is the attempt to argue for the existence of abstract mathematical objects on the basis of the existence of mathematical explanations of empirical phenomena.

The debate on the enhanced indispensability argument is very lively and the literature on it is growing rapidly (Bangu 2014). One point that deserves to be stressed is that nominalists do not reject mathematics – they reject mathematical objects (as Azzouni (2012) and Yablo (2012) correctly note). What needs to be shown in order to establish the existence of abstract mathematical objects on the basis of the existence of explanations of empirical phenomena that appeal to some mathematical theorems is that the truth of those mathematical theorems entails the existence of mathematical objects. Baker (2005) shows that one explanation of some facts concerning the life cycles of cicadas appeal to some number theoretic results. But as Yablo (2012) points out, there are nominalistic interpretations of number theory: the platonist needs to argue that the number theoretic results, interpreted in a nominalistic-friendly way, loose their explanatory power. Until we understand better what it means for mathematics to play an explanatory role, there seems to be

[8] Yablo criticizes the argument "we cannot imagine-without-numbers a complex world [therefore] we cannot imagine a complex world lacking in numbers", Yablo (2012, p. 1014).

a gap between the recognition that mathematical results can have explanatory value (in some sense) and the conclusion that abstract mathematical objects exist (Saatsi 2017: 893). It should also be noted, in connection with the distinction between the representational and the explanatory role of abstract mathematical objects, that sometimes it might be difficult to trace the distinction between the elements of our representation of the world that are mere descriptive aids and those that are genuine elements of the situation that we want to represent. As Quine put it:

> The fundamental-seeming philosophical question, How much of our science is merely contributed by language and how much is a genuine reflection of reality?, is perhaps a spurious question which itself arises wholly from a certain particular type of language. (Quine 1953: 78)

Until now I have insisted on how the question of the existence of abstract objects can be approached following Quine's methodology. Let me end by pointing at different ways in which one can depart from Quine's approach.

Quine's focus on science, in particular theoretical physics, as the source of evidence capable of resolving the dispute over the existence of abstract objects has deeply influenced the contemporary debate and produced a lot of interesting results. But it should be noted that there might be different ways to approach the question whether abstract objects exist, which might be of particular interest for those who focus on abstract objects that are not mathematical objects.

It is true that one could try to make a case for the existence of abstract types (Wetzel 2009) based on their role in our scientific theorizing. And a case for the existence of scientific models conceived as abstract objects would probably be based on an analysis of the scientific practice. But there might be other abstract objects that do not play an essential role in any of the hard sciences. The case for the existence of fictional characters, conceived as abstract objects, is normally based on an analysis of areas of discourse such as literary criticism (Van Inwagen 1977). The case for admitting artworks conceived as abstract objects in our ontology is based on an analysis of art practices and art criticism (Mag Uidhir 2013). And it is natural to assume that the case for entities like propositions would probably come from disciplines like linguistics or the philosophy of language/mind, which are not usually included in the list of the hard sciences.

Apart from broadening the range of areas of discourse to be taken in account, one could depart from Quine's method in more radical ways. One could argue that the existence of abstract objects of various kinds is not a theoretical hypothesis that we should accept in light of its explanatory value, but simply a consequence of uncontroversial truths combined with the rules of usage for certain words. According to the approach defended in Thomasson (2015) ontology should be *easy*: the existence of numbers, for instance, should be acknowledged as a consequence of the fact that (a) I have two hands and (b) if I have two hands, then the number of my hands is two, where (a) is an uncontroversial truth and (b) is a rule that governs the use of the word "number" in English. Similarly, according to neo-Fregeans like Hale and Wright (2003), abstract objects might be introduced by so-called abstraction principles, bi-conditionals which work as implicit definitions of certain concepts:

Hume's principle (the number of the Fs is identical to the number of the Gs if and only if there is a 1-1 relation between the Fs and the Gs), for instance, is presented by the neo-Fregeans an implicit definition of the concept of number. Both the easy ontology approach and the one adopted by the neo-Fregeans bear some connections to Carnap's views on ontology (Carnap 1950), which have recently attracted the attention of several philosophers (see Blatti and Lapointe 2016).

Another aspect of Quine's way of setting the stage for the debate about abstract objects is the idea that we should posit abstract objects only when it is indispensable to do so. As we saw in the previous section, Zalta's theory of abstract objects posit a wealth of abstract objects (one for each group of properties, roughly speaking), including objects which hardly serve any deep theoretical purpose: not only the theory has a place for objects like Leibniz's monads, which arguably are not indispensable to theoretical physics, but also for rather bizarre abstract objects like the one encoding uniquely the property of being either Spanish or a prime number. Quine's view on abstract objects was that we should reluctantly accept them because reference to them is unavoidable. Zalta's view seems rather to be that we have an unproblematic way to know about abstract objects (essentially via the comprehension principle) so that there is no reason not to admit them in our ontology.

Yet another aspect of Quine's methodology that might be challenged is his conviction that it does not make sense to investigate the ontological consequences of our ordinary way of speaking: "ordinary language is slipshod [. . .] a fenced ontology is just not implicit in ordinary language" (Quine 1981, pp. 9–10). Also that assumption has been challenged. Part of the task of natural language ontology (Moltmann 2013, 2017) is the description of the kinds of objects whose existence is presupposed by natural language, using the methods of contemporary linguistics. This analysis might reveal a tension between the opinions of philosophers about what there is and the kind of entities that are presupposed by natural language.

I reviewed some answers to the questions what is an abstract object? Why should we debate over the existence of abstract objects? How should we conduct such a debate? As we are going to see, the present volume contributes to the advancement of the debate on abstract objects by suggesting new ways to address these questions.

References

Armstrong, D. M. (1978). *Universals and scientific realism* (Vol. I and II). Cambridge: Cambridge University Press.

Azzouni, J. (2012). Taking the easy road out of dodge. *Mind, 121*(484), 951–965.

Baker, A. (2005). Are there genuine mathematical explanations of physical phenomena? *Mind, 114*(454), 223–238.

Baker, A. (2009). Mathematical explanation in science. *British Journal for the Philosophy of Science, 60*(3), 611–633.

Baker, A., & Colyvan, M. (2011). Indexing and mathematical explanation. *Philosophia Mathematica, 19*(3), 323–334.

Balaguer, M. (1995). A platonist epistemology. *Synthese, 103*(3), 303–325.

Balaguer, M. (1998). *Platonism and anti-platonism in mathematics*. Oxford and New York: Oxford University Press.

Bangu, S. (2014). *Indispensability of mathematics*. OBO. Retrieved September 29, 2017, from http://www.oxfordbibliographies.com/view/document/obo-9780195396577/obo-9780195396577-0241.xml

Benacerraf, P. (1973). Mathematical truth. *Journal of Philosophy, 70*(19), 661–679.

Blatti, S., & Lapointe, S. (Eds.). (2016). *Ontology after Carnap*. Oxford: Oxford University Press.

Burgess, J. P., & Rosen, G. A. (1997). *A subject with no object: Strategies for nominalistic interpretation of mathematics*. Oxford: Oxford University Press.

Carnap, R. (1950). Empiricism, semantics, and ontology. *Revue Internationale de Philosophie, 4*(11), 20–40.

Colyvan, M. (2002). Mathematics and aesthetic considerations in science. *Mind, 111*(441), 69–74.

Colyvan, M. (2010). There is no easy road to nominalism. *Mind, 119*(474), 285–306.

Contessa, G. (2010). Scientific models and fictional objects. *Synthese, 172*(2), 215–229.

Dorr, C. (2008). There are no abstract objects. In T. Sider, J. Hawthorne, & D. W. Zimmerman (Eds.), *Contemporary debates in metaphysics*. Oxford: Blackwell.

Dorr, C. (2010). Of numbers and electrons. *Proceedings of the Aristotelian Society, 110*(2pt2), 133–181.

Enoch, D. (2010). The epistemological challenge to metanormative realism: How best to understand it, and how to cope with it. *Philosophical Studies, 148*(3), 413–438.

Field, H. (1980). *Science without numbers*. Oxford: Oxford University Press.

Field, H. (1989). *Realism, mathematics & modality*. Oxford: Blackwell.

Fine, K. (2005). Our knowledge of mathematical objects. In T. Z. Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology* (pp. 89–109). Oxford: Clarendon Press.

Goodman, N., & Quine, W. V. (1947). Steps toward a constructive nominalism. *The Journal of Symbolic Logic, 12*(4), 105–122.https://doi.org/10.2307/2266485.

Hale, B., & Wright, C. (2001). *The reason's proper study: Essays towards a neo-Fregean philosophy of mathematics*. Oxford: Oxford University Press.

Knowles, R., & Liggins, D. (2015). Good weasel hunting. *Synthese, 192*(10), 3397–3412.

Lewis, D. (1986). *On the plurality of worlds*. Oxford: Basil Blackwell.

Liggins, D. (2006). Is there a good epistemological argument against platonism? *Analysis, 66*(2), 135–141.

Liggins, D. (2008). Quine, Putnam, and the 'Quine-Putnam' indispensability argument. *Erkenntnis, 68*(1), 113–127.

Liggins, D. (2010). Epistemological objections to platonism. *Philosophy Compass, 5*(1), 67–77.

Liggins, D. (2016). Grounding and the indispensability argument. *Synthese, 193*(2), 531–548.

Linnebo, Ø. (2018). Platonism in the philosophy of mathematics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2018 Edition). URL = https://plato.stanford.edu/archives/spr2018/entries/platonism-mathematics/

Linsky, B., & Zalta, E. (1995). Naturalized platonism versus platonized naturalism. *The Journal of Philosophy, 92*(10), 525–555. https://doi.org/10.2307/2940786.

Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. Harmondsworth: Penguin.

Mag Uidhir, C. (Ed.). (2013). *Art & abstract objects*. Oxford: Oxford University Press.

Malament, D. (1982). Science without numbers by Hartry H. Field [Book Review]. *Journal of Philosophy, 79*(9), 523–534.

Melia, J. (2000). Weaseling away the indispensability argument. *Mind, 109*(435), 455–480.

Melia, J. (2008). 5. A world of concrete particulars. *Oxford Studies in Metaphysics, 4*, 99.

Moltmann, F. (2013). *Abstract objects and the semantics of natural language*. Oxford: Oxford University Press.

Moltmann, F. (2017). *Natural language ontology. Oxford encyclopedia of linguistics*. Oxford: Blackwell.

Morrison, M. (2007). Spin: All is not what it seems. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, 38*, 529–557.

Plebani, M. (2018). The indispensability argument and the nature of mathematical objects. *Theoria: An International Journal for Theory, History and Fundations of Science, 33*(2), 249–263.

Plebani, M. (MS). *The refutation of nominalism*! https://www.academia.edu/33566323/The_Refutation_of_Nominalism.docx

Putnam, H. (2012). Indispensability arguments in the philosophy of mathematics. In M. DeCaro & D. Macarthur (Eds.), *Philosophy in an age of science: Physics, mathematics, and skepticism* (pp. 181–201). Harvard: Harvard University Press.

Quine, W. V. O. (1953). *From a logical point of view*. Cambridge, MA: Harvard U.P.

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Quine, W. V. O. (1969). *Ontological relativity and other essays*. New York: Columbia U.P.

Quine, W. V. O. (1981). *Theories and things*. Cambridge, MA: Belknap Press.

Quine, W. V. O. (1986). Reply to Chanles Persons. In L. Hahn, & P. Schilpp (Eds.), *The philosophy of W. V., Quine* (pp. 396–403). La Salle, IL: Open court.

Rosen, G. (1994). Objectivity and modern idealism: What is the question? In J. O'Leary-Hawthorne & M. Michael (Eds.), *Philosophy in mind* (pp. 277–319). Dordrecht: Kluwer Academic Publishers.

Rosen, G. (2017). Abstract objects. In: E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2017 Edition), forthcoming URL = https://plato.stanford.edu/archives/win2017/entries/abstract-objects/

Saatsi, J. (2017). Dynamical systems theory and explanatory indispensability. *Philosophy of Science, 84*(5), 892–904.

Szabó, Z. G. (2003). Nominalism. In M. J. Loux & D. W. Zimmerman (Eds.), *The Oxford handbook of metaphysics*. Oxford: Oxford University Press.

Thomasson, A. (1999). *Fiction and metaphysics*. Cambridge: Cambridge University Press.

Thomasson, A. L. (2015). *Ontology made easy*. New York: Oxford University Press.

Van Inwagen, P. (1977). Creatures of fiction. *American Philosophical Quarterly, 14*(4), 299–308.

Wetzel, L. (2009). *Types and tokens: On abstract objects*. Cambridge, MA: MIT Press.

Williamson, T. (2013). *Modal logic as metaphysics*. Oxford: Oxford University Press.

Yablo, S. (2005). The myth of seven. In M. E. Kalderon (Ed.), *Fictionalism in metaphysics* (pp. 88–115). Oxford: Clarendon Press.

Yablo, S. (2010). *Things: Papers on objects, events, and properties*. Oxford: Oxford University Press.

Yablo, S. (2012). Explanation, extrapolation, and existence. *Mind, 121*(484), 1007–1029.

Zalta, E. (2016). *Principia metaphysica*. http://mally.stanford.edu/principia.pdf. (Version of October 28, 2016).

Part I
Enhanced Indispensability and Type Theories

Chapter 2
Purely Physical Explananda: Bistability in Perception

Sam Baron

Abstract Call an explanation of an empirical fact that appeals indispensably to mathematical facts: an extra-mathematical explanation. There has been a recent rush of attempts to establish the existence of such explanations. One of the challenges in doing so, however, is to find a case of explanation in which the explanandum is genuinely free of mathematics. In this chapter I first outline this challenge before drawing attention to an important class of explanations that seem to meet the challenge head on. The explanations in question are explanations that appeal to bistability: a situation in which the differential equations modelling a system have two roots, and thus the physical system has two stable states, but is unstable between them. I focus, in particular, on the case of perceptual bistability in which an ambiguous image of a duck-rabbit seems to 'switch' back and forth between the two disambiguations (duck and rabbit).

2.1 Introduction

An extra-mathematical explanation is the explanation of a physical fact that appeals, indispensably, to mathematical facts. Explanations of this kind are important for two broad reasons. First, if there are genuine extra-mathematical explanations, then theories of scientific explanation will need to be generalised so as to handle explanations of this kind. Theories of scientific explanation have been historically bad at managing extra-mathematical explanations, tending to focus, instead, on causal explanations. Because of this tendency, generalising available theories of scientific explanation in the required manner is not straightforward. Extra-mathematical explanations may not be causal explanations, and so the machinery that has been developed to model scientific explanations may need to be substantially re-tooled.

S. Baron (✉)
University of Western Australia, Crawley, WA, Australia
e-mail: samuel.baron@uwa.edu.au

J. L. Falguera, C. Martínez-Vidal (eds.), *Abstract Objects*, Synthese Library 422,
https://doi.org/10.1007/978-3-030-38242-1_2

Second, extra-mathematical explanations are important in arguments for realism about mathematical objects. The standard argument in favour of the existence of mathematical objects is the indispensability argument. Very roughly, the indispensability argument aims to establish the existence of mathematical objects via the role that mathematics plays in science. The basic thought is that we should believe in the existence of anything that is indispensable to our best scientific theories. Mathematical objects, it has been argued, are indispensable to those theories and so it follows that we should believe in their existence. It is, however, generally agreed that indispensability is not sufficient to establish the existence of mathematical objects. Rather, mathematical objects must be indispensable in the *right way* in order to reasonably attract a realist attitude. In particular, it must be shown that mathematical objects are indispensable to our best scientific explanations of physical phenomena. If this can be shown, then the existence of mathematical objects can be established via inference to the best explanation (IBE) in a manner that is directly analogous to the way in which the existence of unobservable entities – like blackholes – gets established.

My interest in this paper is on the second motivation for taking extra-mathematical explanations seriously. In order to establish the existence of mathematical objects via IBE it must be shown that there are genuine cases of extra-mathematical explanation. That is, cases in which mathematics plays an indispensable role in the explanation of physical phenomena. A range of such cases have been identified. One of the difficulties with these cases, however, is that they typically feature mathematics in the explanandum. This is a problem because the presence of mathematics in the explanandum renders the use of these explanations question begging in arguments for realism.

My goal in this chapter is to draw attention to extra-mathematical explanations in which the explananda are not mathematical in nature. Such cases, I will suggest, provide a better basis upon which to argue for mathematical realism. I will begin by outlining an argument in favour of mathematical objects based on IBE, before considering an example of extra-mathematical explanation: Baker's case of the North American cicadas. I will then use this case to demonstrate the problem with mathematical explananda for realism about mathematical objects. After that, I will draw attention to a new class of explanations involving mathematics. I will look in detail at one case in particular: the case of perceptual bistability in which an ambiguous image (such as an image of a duck-rabbit) seems to 'switch' back and forth between two disambiguations (duck and rabbit).

2.2 The Cicadas

Consider the enhanced indispensability argument (Baker 2009, p. 613):

The Enhanced Indispensability Argument

[1] We ought rationally to believe in the existence of any entity that plays an indispensable explanatory role in our best scientific theories.
[2] Mathematical objects play an indispensable explanatory role in science.
[3] Hence, we ought rationally to believe in the existence of mathematical objects.

The enhanced indispensability argument seeks to establish the existence of mathematical objects via IBE. The idea being that mathematical entities get caught up in confirmation via IBE in the same way that any unobservable entity does in science. Thus, we should believe in the existence of mathematical entities for the same reason that we should believe in the existence of black holes. Accordingly, the argument is aimed primarily at scientific realists. Scientific realists uphold a commitment to the viability of IBE as a form of inference. Scientific anti-realists, by contrast, often deny that IBE is an acceptable form of inference. The basic idea behind the enhanced indispensability argument, then, is to push scientific realists toward realism about mathematical objects. This pressure is uncomfortable for many scientific realists who find the existence of mathematical entities to be unpallateable for one reason or another.

The viability of the enhanced indispensability argument turns on finding genuine cases of extra-mathematical explanation: explanations in which mathematical objects play an indispensable explanatory role. Not just any such case will do, however. We must be careful to only use cases in which there is no mathematics in the explanandum. To see why, it is useful to consider a particular example. To this end, let us take a brief look at what has become the poster-child for extra-mathematical explanation: Baker's (2005) case of the North American cicadas.

In North America there are two sub-species of magicicada that have prime-numbered, periodical life-cycles of 13 and 17 years respectively. Possession of a periodical life-cycle means that the magicicada spends most of its time in the ground, in its larval form. Every 13 or 17 years, depending on the sub-species, the magicicada arises in its adult form for a period of around two weeks during which it eats, breeds, dies and repeats the cycle. The phenomenon of magicicada swarming gives rise to an important explanatory question that biologists have sought to answer, namely: why are the life-cycles of the North American magicicada prime-numbered? The answer lies in number theory: because of the manner in which prime-numbers factorise, having a prime-numbered life-cycle turns out to be the optimal strategy for avoiding predation from predators that also have periodical life-cycles. For instance, compare 13 to 12. 12 has factors 2, 3, 4, 6, 12 and 1. So a cicada with a 12-year life-cycle will overlap with predators that have periodical life-cycles of 2, 3, 4, 6 and 12. 13, by contrast, only has factors 13 and 1, and so will overlap only with predators that have 13-year life-cycles. This generalises to all other numbers in the 12–18-year range: 13 and 17 outshoot all of the alternatives when it comes to predator avoidance.

As Baker (2005, 2009) argues, mathematical facts about primes appear to be playing an indispensable role in the explanation for cicada life-cycle length. In particular, the periodical intersection of prime life cycles can be captured within a general mathematical framework that describes unit cycles: abstract mathematical cycling phenomena (Baker 2017). By focusing on unit cycles and on prime periodicity it is possible to demonstrate the optimality of having a prime-numbered life-cycle length with respect to predator avoidance.[1]

Notice, however, that the explanandum in the cicada case is, itself, mathematical. The fact that we are trying to explain is the fact that cicadas possess prime life-cycle lengths. Bangu (2008) argues that the mathematical nature of the explanandum poses a substantial problem for the use of the cicada case in the context of an argument in favour of mathematical realism. The trouble comes this way. If the explanandum of the cicada case is itself a mathematical fact, then in order for the explanatory project to get started we must presuppose the truth of a mathematical fact. Any argument in favour of mathematical facts on the basis of the cicada case would therefore seem to be question begging. The problem was first recognised by Leng (2005, p. 174), who writes:

> The reason that this [cicada] argument can't be used is that, in the context of an argument for realism about mathematics, it is question begging. For we also assume here that genuine explanations must have a true explanandum, and when the explanandum is mathematical, its truth will also be in question.

Most of the examples of extra-mathematical explanation offered to date face a similar problem. These explanations tend to feature mathematics in the explanandum. So the problem that Leng and Bangu raise appears to be a general problem for the mathematical realist.

Baker (2009, pp. 221–222) suggests that in order to address the problem, cases of extra-mathematical explanation should be modelled via the following general argument pattern, which combines IBE with a kind of bootstrapping:

[1] Data, D;
[2] Tentative hypothesis, H;
[3] Explanation, E, of H, which also can be extended to yield an explanation, E*, of D;
[4] E* is the best explanation of D;
[5] Hence we ought to believe E*, and thereby E;
[6] But D and E together imply H;
[7] Hence we ought to believe H.

When applying this argument pattern to a case of mathematical explanation, the crucial step is to take the explanandum – which is mathematical in nature – and treat it as a tentative hypothesis, rather than something we take to be true. To understand Baker's solution, let us work through the cicada case. First, we formulate the explanandum as a tentative hypothesis, or set of hypotheses. This gives us:

[1] Other examples of extra-mathematical explanation include explanations for: the shape of honeycomb cells (Lyon and Colyvan 2008), the structure of seeds in a flower (Lyon 2012), the search patterns of fully-aquatic marine predators (Baron 2014), the minimal shape of Plateau's soap film (Lyon 2012) the location of the Kirkwood gaps (Colyvan 2010), the Fitzgerald-Lorenz contraction of bodies at relativistic speeds (Colyvan 2002), the instability of high-energy galaxies (Lyon and Colyvan 2008), the use of real-valued functions to understand physical systems (Peressini 1997), and the use of Hilbert spaces in quantum mechanics as a basis for explaining quantum phenomena (Steiner 1995 and Peressini 1997).

(1*) The length (in years) of the life cycle of cicada subspecies A is prime.
(2*) The length (in years) of the life cycle of cicada subspecies B is prime.

Next, we identify two pieces of data that would be explained by the same explanation that explains either (1*) or (2*). Baker suggests the following two facts:

(1) The length (in years) of the life cycle of cicada subspecies A is 13.
(2) The length (in years) of the life cycle of cicada subspecies B is 17.

Note that while these two claims contain mathematics, their truth does not presuppose mathematical realism. That's because both claims can be fully expressed in the language of first-order predicate logic with identity. There is no sense in which the use of numbers is essential to the relevant claims. The same is not true for (1*) and (2*). The concept of primeness that appears in (1*) and (2*) cannot be translated in the same way.

Next, the extra-mathematical explanation of cicada life-cycle length is offered as an explanation of (1*) and (2*), which can also be extended to yield an explanation of (1) and (2) (given that (1) and (2) are specific instances of (1*) and (2*)). Because the extra-mathematical explanation of cicada life-cycle length is the best explanation of (1*) and (2*) we should believe it. Finally, we should believe (1*) and (2*) because they are implied by two other things that we believe: namely the extra-mathematical explanation of their truth together with (1) and (2). The result is that we should believe that the extra-mathematical explanation of cicada life-cycle length is true, and we should also believe that (1*) and (2*) are true.

As Baker notes, this procedure does not beg the question. One need not presuppose the truth of any mathematical claim at any point. Rather, the explanandum and the explanans get confirmed together, in the same inferential process. This is presumably why Baker describes the procedure as a bootstrapping process. We establish the truth of the explanandum via the explanans, while simultaneously establishing the truth of the explanans via the explanandum. Note that the inferential procedure is not circular. It is not as though we *first* establish the explanans and then use that to establish the explanandum, after which we then use the explanandum to establish the explanans. Rather, the two facts – explanans and explanandum – are mutually supportive.

Baker's proposal is intriguing, though it is somewhat unclear how general the solution is. It is fortunate that, in the cicada case, it is possible to find two pieces of data that are entailed by the core explanandum, and that can be expressed in purely logical terms. That this is possible has a lot to do with the mathematics that features in the case. The mathematics being used in the cicada case is, primarily, number theory, and the mathematical entities at issue are numbers. We have well-known resources available for translating basic number talk into logic, which makes claims (1) and (2) available in a non-question-begging argument for mathematical realism. In other cases of extra-mathematical explanation, the same resources may not be available. That's because the mathematics at issue in other cases departs substantially from number theory, and goes well beyond the realms of mathematics that can be expressed in purely logical terms.

In the end, though, the generalisability of Baker's solution is not all that important. All it takes to get the enhanced indispensability argument off the ground is a single case of extra-mathematical explanation in science. So even if Baker's solution works only in the cicada case, that is still enough to address the problem that Leng and Bangu have identified. Of course, the enhanced indispensability argument grows stronger as more examples of extra-mathematical explanation are brought to light, and so the generalisability of Baker's solution is desirable, but ultimately not essential for establishing mathematical realism.

There is, however, a deeper problem with Baker's solution. As indicated, the enhanced indispensability argument seeks to establish the existence of mathematical objects in the same way that a scientific realist establishes the existence of unobservables, via IBE. The goal of the argument is to push the scientific realist to broaden her realism. The argument itself, then, is only as strong as the analogy between the use of IBE to establish the existence of unobservables, and the use of IBE to establish the existence of mathematical objects. The analogy is at its strongest when the two instances of IBE – the one used to establish black holes and the one used to establish mathematical objects – operate in the same manner. If the two notions of IBE do not operate in the same manner, then an ancillary argument is required; one must argue that the form of IBE used to establish mathematical objects is a form of IBE that a scientific realist ought to accept, given her broader realist leanings.

The form of IBE discussed above – the one that Baker suggests as a way of addressing Leng's objection – is unlike standard forms of IBE in science. Standard IBE does not have the bootstrapping character that Baker's IBE does. This leaves an opening for a scientific realist who wishes to resist the enhanced indispensability argument. She can argue that the only type of IBE that should attract a realist commitment is one that does not involve bootstrapping the explanandum. Moreover, such a position is *prima facie* plausible. There is something inferentially odd about Baker's bootstrapping IBE. When we use standard IBE, we are compelled to accept the truth of the explanans because it is required to explain something that we take to be true. In the case of Baker's bootstrapping IBE, we do not take the explanandum to be true. Rather, the truth of the explanandum gets established via IBE. While it seems relatively obvious that a realist attitude is justified when we are trying to explain something that we assume to be true, it is less obvious that realism is justified when we are trying to explain something that we don't take to be true, even if the truth of that fact ultimately gets confirmed via the explanatory procedure.

Note that I don't take this worry to be a decisive objection against Baker's solution. I do, however, believe it motivates looking for an alternative. In particular, it would be preferable to preserve the analogy between the case of unobservables and the case of mathematical objects by keeping the inferential structure of IBE fixed across the two cases. One way to do this is to find a case of extra-mathematical explanation in which the explanandum is purely physical. My goal in what remains is to outline such a case.

2.3 Bistability

The case of extra-mathematical explanation that I will focus on makes use of bistability. I will briefly introduce the mathematical notion of bistability, before looking in detail at a particular case of bistability that arises in perception.

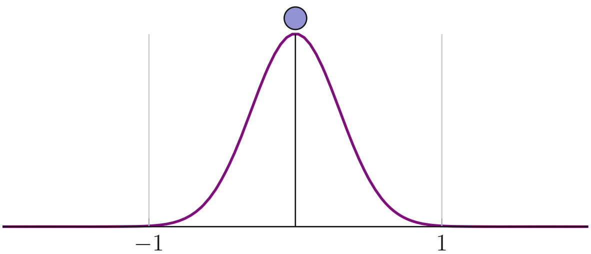
Roughly put, a system is bistable when it has three states, two of which are stable and one of which is unstable. A very simple example of a bistable system is a light switch. A light switch has three states: on, off and in-between. Light switches are designed to be stable in the 'on' position and stable in the 'off' position, but to be unstable in the 'in-between' position. Thus, for a properly designed light switch, when you try to leave the switch in a position that is between off and on, the switch will be unstable and will tend toward stability by flicking on or off.

Here's another example of a simple bistable system. Consider a peak of the kind depicted in Fig. 2.1. Imagine that a ball is dropped at the top of the peak. The top of the peak is an unstable state of the system, in this sense: when the ball is placed at the peak, it won't stay there. It must fall to one of the two sides. The bottom of the trough on the left of the peak, and the bottom of the trough on the right of the peak each correspond to stable states. When the ball reaches one of those stable states it stays there, unless it is picked up and placed back at the peak of the system.

Differential equations describing any such system tell us how the system will evolve, based on the initial conditions. For example, if we let the peak of the system correspond to 0, then we can formulate a differential equation with two solutions depending on the starting point. If the starting point of the system is positively skewed – if the ball is placed at the peak, but slightly to the right – then the solution of the equation will be 1. This corresponds to the base of the trough on the right hand side. If, by contrast, the starting point of the system is negatively skewed – if the ball is placed at the peak, but slightly to the left – then the solution of the equation will be -1. This corresponds to the base of the trough on the left hand side. The differential equation is bistable in so far as it is stable between these two options: the system will tend to end up in one or the other of these two states, based on a choice of initial conditions.

Bistability constitutes the backbone of a range of explanations of physical phenomena. Explanations that appeal to bistability in some manner are good places to look for extra-mathematical explanations. That's because the core concept of

Fig. 2.1 A ball rolling toward one of two end states



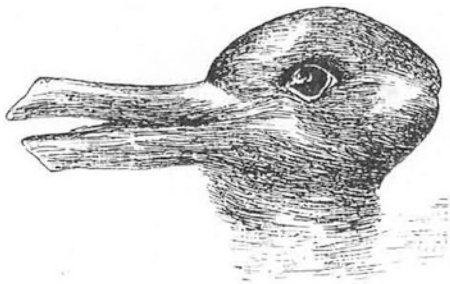


Fig. 2.2 The duck-rabbit is an ambiguous figure that switches between two stable percepts

bistability is standardly encoded using differential equations in roughly the manner just described for the ball case (though the complexity of the differential equations obviously increases with the complexity of the system being modelled).

There is one case, in particular, that is quite striking: the case of bistability in perception. Consider the infamous duck-rabbit illusion (Fig. 2.2).

Human perception of this ambiguous image 'switches' between two unambiguous images: a picture of a duck and a picture of a rabbit. We cannot see both of the images at the same time. However, we must see one. There are a number of models of this switching behaviour. A common aspect of these scientific models of perception is that they use differential equations to model the bistability of the perceptual system.

Very roughly, the perceptual system is treated in a manner that is akin to the ball case sketched above. There are two neural populations: one that corresponds to the perception of the duck and one that corresponds to the perception of the rabbit. These two populations 'compete' for dominance. The perceptual system overall has three states, two of which are stable. One stable state is associated with the dominance of the neural population corresponding to the perception of a duck, the second stable state is associated with the dominance of the neural population corresponding to the perception of a rabbit. The third state, which is unstable, corresponds to equidominance between the two neural populations. The system is bistable, which means that it must always be in one of the two stable states. The third state – the equidominance state – will force the system into one of the two percepts. What this means, in effect, is that you must always see a duck or a rabbit, but never the two at once. There is no stable duck+rabbit percept.

There are two broad neural models of why the switching between the duck and rabbit percepts occurs (Moreno-Bote et al. 2013, pp. 7–8). Both models focus on the firing rates of the two neural populations. Each stable percept (duck versus rabbit) corresponds to the dominance in terms of firing rate of one neural population over the other. The two models differ in the explanations that they provide of why perception switches from a duck to a rabbit and back again repeatedly.

According to the *oscillator model*, the two neural populations sit in a see-sawing relationship with regard to their firing rate (Shpiro et al. 2007a,b, 2009). To see the idea, call the neural population corresponding to the duck percept: A, and the neural population corresponding to the rabbit percept B. As A's firing rate increases, B's decreases, which leads to the dominance of A over B. Over time, however, there is adaptation in the neural population A, which slows A's firing rate.[2] As A's firing rate slows, B's increases, which leads to the dominance of B over A. The same adaptation process that slows the firing rate of population A, however, then kicks in for B, which subsequently slows its firing rate while A's increases, leading to a dominance of A once again. In short, over time, the percepts switch because our neural architecture oscillates between the two dominant states. The periodicity of the oscillation ought to be regular, given the deterministic nature of the oscillator models. Which means that the length of time for which a percept is dominant – the mean dominance duration – ought to be the same for each percept and in each round of oscillation. Because experimental results don't show regular periodicity, an extra stochastic element is typically added into the dynamics of the system to make the model more realistic. The mathematical basis of the oscillator model includes coupled differential equations for the firing rate of the two neural populations A and B (see (2.1) and (2.2)).[3]

$$\tau \frac{d}{dt} r_A = -r_A + f(\alpha r_A - \beta r_B + I_A - a_A + n_A) \tag{2.1}$$

$$\tau \frac{d}{dt} r_B = -r_B + f(\alpha r_B - \beta r_A + I_B - a_B + n_B) \tag{2.2}$$

The second type of model is the *attractor model* (Moreno-Bote et al. 2007). In the oscillator model, the system 'bounces' between two solutions to the associated differential equations. The dynamics of the system is such that it will never 'settle down' into one of those solutions if left on its own. Thus, it will never be the case that one will come to see a duck, and then forever more see only ducks when looking at ambiguous duck-rabbit images. In an attractor model, by contrast, the system will settle down into one of the solutions: those solutions are not just stable, they are fixed points. So, in an attractor model, the perceptual system will settle on a duck or a rabbit and then stay there. This is, however, only true for an isolated system. In a system with noise, there is a random process that forces the system out of one state and back into the other. In this situation, the stochastic nature of the mean duration dominance of a given percept is written into the dynamics of the equation and plays a central role in why the system switches. In the oscillator models, the noise parameter doesn't enter into the explanation of switching; it is only there to explain why there is variation in the mean dominance duration of each percept, following a switch.

The primary attractor model is the double-well potential energy model. In the double-well potential energy model, the dynamics of the two neural populations is

[2] Adaptation corresponds to the decreasing sensitivity of a neural population to a give stimulus. E.g., a bad smell, when first encountered, is hard to take, but it quickly becomes less noticeable as the neural system adapts to the stimulus and slows its firing rate in response.

[3] τ = 10 ms, which is a time constant, f(x) determines the firing rate given energy inputs from a population; I_A and I_B are currents measuring the stimulus strength of the two percepts.

modelled using an energy landscape. The two percepts – the percept of the duck and the percept of the rabbit – correspond to two potential energy wells in this landscape. These wells correspond to the potential energy of the neural populations corresponding to those percepts. As a neural population increases its rate of firing, it uses its potential energy, and the potential energy well for that neural population deepens. The deepening of the energy well for a given neural population increases the mean duration dominance of a percept. In short: as a neural population increases its rate of firing, the mean time that we experience the percept corresponding to that neural population increases.

The two neural populations are related in a see-sawing of potential energy. As the energy well for one neural population increases, the energy well for the other neural population decreases until there is, effectively, no second stable energy state for the system to be in. This is why the system will stay in one of the two states indefinitely: because its presence in one of the states effectively makes the other stable state unreachable, and thus prevents the other neural population from achieving dominance with respect to rate firing. Of course, actual perceptual systems don't just get to duck and stay there: eventually there is always rabbit. To handle this, a noise element is added: an energy spike in the system deepens the energy well of the second neural population, and pushes the dynamics to a tipping point, whereby the energy well for the second neural population makes the stable state for the first population unreachable. This stays the case until another energy spike occurs and so on, creating the switching. The mathematical essence of the attractor model is stated in (2.3).[4]

$$\tau \frac{d}{dt} r = -4r(r^2 - 1) + I_A - I_B + n(t) \tag{2.3}$$

There is a great deal more to say about the difference between the oscillator and attractor models. For present purposes, however, the similarity between the two models is more important than any differences. Both models make use of differential equations to model the bistability of the perceptual system. To be sure, the two models use different differential equations for this purpose, but in a general sense they share a mathematical core. More importantly, the shared mathematical core of the two models is indispensable to the explanations that they provide. In order for the mathematics to be indispensable, it must be the case (roughly) that there is no alternative explanation of the phenomena that possesses the same explanatory power, and that makes no use of the relevant mathematical core. This is indeed the case. While the explanatory approach under consideration has two different forms–oscillator and attractor–in both cases the mathematics is needed to explain the switching phenomena that we observe. There is no model of bistability that does not make use of one of the three differential equations identified above.

[4] $\tau = 10\,\text{ms}$, which is the timescale of the system; I_A and I_B are currents measuring stimulus strength of the two percepts, $n(t)$ is a noise term.

Notice, moreover, that the explanandum in the bistability case is not mathematical. The explanandum is a purely physical phenomenon: the fact that human perceptual systems switch between two distinct percepts of an ambiguous image. Indeed, it is quite implausible to suppose in the case at hand that the mathematics doing the explaining appears anywhere in the explanandum. To see why, suppose, for a moment, that mathematical facts about bistability were to appear in the explanandum in a manner that is analogous to the use of primeness in Baker's cicada case. In this situation, we would end up trying to explain why a given differential equation is bistable (since that is what occurs in the explanandum). But that is just not the explanatory project at hand. What we want to know is why the perceptual system switches, an explanatory question that says nothing about the bistability of any set of equations.

2.4 Carrying the Explanatory Load

So far I have suggested that the case of perceptual bistability is a case of explanation in which mathematics is indispensable and the explanandum–perceptual switching–is a purely physical phenomenon. In recent work on mathematical realism, however, it has become clear that being indispensable to an explanation may not be sufficient to warrant realism about mathematics (see Saatsi 2016). The trouble has to do with the representational capacity of mathematics. Mathematics can and often is used to represent a range of physical facts. This is true for science quite generally, but for many scientific explanations in particular. When we see that a piece of mathematics is explanatorily indispensable to a scientific explanation, we know that there is a particular representation–in the bistability case, it is a scientific model–that contains mathematics, where the mathematics increases the explanatory power of the representation. However, what we don't know is what role the mathematics really plays in explaining the phenomenon being modeled. While the increase in explanatory power may be because the mathematics enables the representation to more accurately represent the mathematical facts that explain a particular physical phenomenon, it could also be because the mathematics enables the representation to more accurately represent some physical fact that is really doing the explanatory work.

It seems plausible that the explanatory indispensability of mathematics only warrants mathematical realism if we can rule it out that the mathematics is merely representing some underlying physical fact that does all of the explanatory work. The question, then, is whether we can rule it out that the mathematics is playing a merely representational role in the bistability explanation offered above. One might think not: while mathematical facts are needed to build models of perceptual bistability, those facts do not carry the explanatory load in the models in question. Rather, what carries the explanatory load in each of the models at issue is some purely physical fact.

For example, consider the oscillator model of perceptual bistability. According to this model, perception switches between a duck percept and a rabbit percept due to adaptation within the system, which is a purely physical feature of human neural architecture. It is because our neural system undergoes a process of adaptation with regard to a particular percept (say, of the duck) that the neural correlates of that percept slow their rate of firing. Switching then occurs. The differential equations are used to model the states, and indeed the switching process itself. But the bulk of the explanatory work appears to be done by the adaptation process; it is unclear what, if anything, the mathematics adds to the explanation, over and above an ability to state the explanation in a formally precise fashion.

Similar considerations apply to the attractor model. As already indicated, the attractor model explains the switching between a duck percept and a rabbit percept in terms of noise entering the system. Energy spikes occur randomly within the perceptual system which force the system out of the potential energy well corresponding to one neural population and into the potential energy well corresponding to the other neural population. As with the oscillator model, it is unclear what work the mathematics is actually doing in explaining why it is that switching occurs. Granted, we need a set of differential equations to model the dynamics of the system broadly construed, but the explanation for why human perception flicks back and forth between a duck and a rabbit appears to be carried primarily by the noise element that disrupts the system.

One way to try and address the concern that mathematics is merely representational in the attractor and oscillator models is to consider a different ‘why’ question. In particular, we can focus instead on a ‘why’ question in the neighbourhood that essentially controls for the adaptation process and the presence of noise in the system. The idea being that we may be able to rule it out that these physical facts explain the phenomenon of interest, leaving just the mathematical parts of the explanation to pick up the explanatory load. So, for example, rather than asking: why does perception switch between a duck percept and a rabbit percept? Let us, instead, compare the perceptual switching that occurs in the duck-rabbit case, with the perception of a non-ambiguous figure: a picture of a rabbit or a picture of a duck (See Fig. 2.3).

Now, consider the following why question: why is there perceptual switching in Fig. 2.2 but no perceptual switching in Fig. 2.3? Notice that the explanation for the difference cannot appeal either to neural adaptation or to noise. In both cases, we are faced with neural adaptation of the percept. When we are seeing a duck in Fig. 2.2, and when we are seeing a duck in Fig. 2.3, adaptation occurs in the neural populations that correspond to the two percepts. Similarly, in both cases the same random noise that triggers switching in the attractor model is present in both cases. Human perceptual systems will enjoy energy spikes when looking at ambigious figures such as Fig. 2.2, just as much as when looking at unambiguous figures such as Fig. 2.3. Finally, notice that very similar processes are involved in the two kinds of percept. Indeed, it is plausible to suppose that the same neural population that corresponds to the duck percept in Fig. 2.2 also corresponds to the duck percept in Fig. 2.3. This same neural population is firing in both cases, at some rate, which is

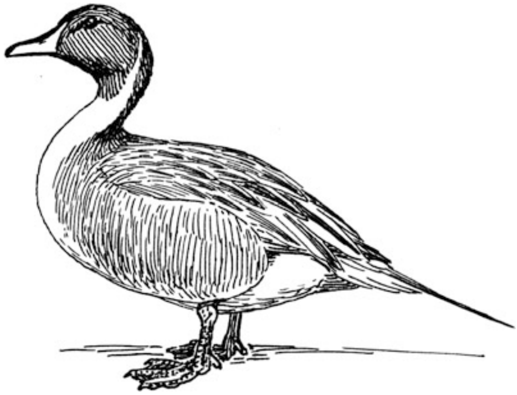


Fig. 2.3 An unambiguous picture of a duck

why we experience what we do. If noise and adaptation can't do any explanatory work, and the neural processes that are involved in the two perceptual cases are more or less the same, then the one remaining element appears to be the mathematical fact that the dynamical equations are bistable, which enables the physical system to behave in a certain manner.

The difference between the two cases thus potentially comes down to the dynamics of the two neural systems. With respect to the neural process corresponding to Fig. 2.3, there is only one stable state for the dynamics of the system to enter into. This is the stable state corresponding to a duck percept. With respect to the neural process corresponding to Fig. 2.2, there are two stable states for the dynamics of the system to enter into. These correspond to the duck and rabbit percepts respectively. The reason why the dynamics for a system has two stable states and not one appears to be mathematical in nature. When the differential equations for a system have just one solution, the system has only one option for stability. When the differential equations for a system have two solutions, by contrast, those equations make available a second stable option for the system. It is, potentially, because the dynamics of the system is constrained by a different set of equations with distinct mathematical properties that perceptual switching is an available option and is thus capable of occurring.

But is that right? It may be that stability in a physical system is itself a physical property, one that the differential equations are merely representing. Perhaps it is just a lawful feature of the system that it is stable in some cases and unstable in others. In short, the constraint that the system is under may not be a mathematical one. It may be that the differential equations for the system are bistable because of the physical stability properties of the system rather than vice versa, as I have been suggesting. There is also the fact that Figs. 2.2 and 2.3 are different, and it may be that the physical features of the figures ultimately explains the difference in switching between the two cases.

What we really need is a test for explanatory relevance: a way to determine when mathematics itself is carrying the explanatory load, rather than merely representing some physical feature that is doing this work. Developing such a test is a large project, and I cannot hope to complete it here. What I will do, however, is offer an example of the kind of test that might do the job, as a basis for future work.

Let us start by saying a bit more about what it is for something to be carrying the explanatory load. There are many ways we could try to cash this out, but for present purposes I will focus on a difference-making account. Thus, I will assume that something is carrying the explanatory load in an explanation if it makes a difference to the expanandum.

There are various ways to test for difference-making, but the one I will use here is a counterfactual test. Roughly, for any A and B, if it is true that had A not been the case, B would not have been the case, then we have a reason to suppose that A makes a difference to B. Note that this is not supposed to be an analysis of difference-making. The thought, rather, is that counterfactuals provide a way to probe for information about what is, and what is not, making a difference to some explanandum.

Consider the following simple example. Suppose that we explain the breaking of the window in terms of the fact that Suzy threw a rock at it. The fact that Suzy threw a rock at the window carries the explanatory load in this explanation. It carries the explanatory load, because the throwing of the rock makes a difference to whether or not the window broke. How do we know? Because of the counterfactual structure of the situation. If Suzy had not thrown the rock, then the window would not have broken, and so we have good reason to suppose that her throwing is a difference-maker.

Of course, not every aspect of an explanation carries the explanatory load in this sense. For example, suppose that, in formulating the explanation for why the window broke, we specify that the rock Suzy threw was of a particular type. Thus, we say that Suzy threw an igneous rock at the window. While this may be true, the fact that the rock is an igneous rock does not carry any of the explanatory load in this explanation. We can also work this out by looking at the following counterfactual: If Suzy had not thrown an igneous rock at the window, it would not have broken. This counterfactual is false: a metamorphic rock would have done just as well.

Now, consider the first explanation discussed above, the explanation of why we experience perceptual switching between a duck and a rabbit in Fig. 2.2. Suppose that the oscillator model is the correct explanation for why it is that we experience switching in this case. As discussed, one of the key aspects of the oscillator model is adaptation. Switching occurs, according to the oscillator model, because a process of adaptation occurs for a given percept, which slows the rate of neural firing for one neural population, which enables the other neural population to achieve dominance. We have reason to believe that the adaptation process helps to carry the explanatory load because the following counterfactual is true: if the neural system involved in a case of ambiguous perception of a duck-rabbit had not undergone adaptation, it would not have undergone switching between percepts of a duck and a rabbit. This counterfactual is true because, were we to remove the adaptation process, the chief mechanism behind the switching between the duck and rabbit percepts would no longer be available to trigger any switching process.

Alternatively, suppose that the attractor model is the correct explanation for perceptual switching between a duck and a rabbit in Fig. 2.2. As discussed, a central component of this model is a noise element. Random energy spikes in the neural system ultimately induce the switching from a duck percept to a rabbit percept. We have reason to believe that the noise component is helping to carry the explanatory load, because the following counterfactual is true: if a neural system involved in a case of ambiguous perception of a duck-rabbit featured no noise, it would not have undergone switching between percepts of a duck and a rabbit. That the counterfactual is true is something that can be seen from the dynamics of the system. As already discussed, the double-well potential energy model is such that when the potential energy well for one of the percepts deepens, the other percept effectively becomes out of reach for the system until an energy spike occurs. Without the energy spike, switching cannot happen.

So far, then, the analysis of 'carrying the explanatory load' appears to yield the right results for the first case of perceptual bistability – the perceptual switching associated with an ambiguous figure. This gives us some measure of confidence in the counterfactual test. We can now apply the test to the mathematical facts that occur in the second case of perceptual bistability – the difference in perception between Figs. 2.2 and 2.3. In order for the mathematics to be carrying the explanatory load in this case, the mathematical facts need to be difference-makers for the phenomenon of interest. As such, there are two counterfactuals of interest (see below). Both counterfactuals link the mathematical features of the relevant explanation with the explanandum.

> [CF_1] If the differential equations governing the neural system corresponding to the perception of Fig. 2.2 had not been bistable, Figs. 2.2 and 2.3 would have both lacked perceptual switching.

> [CF_2] If the differential equations governing the neural system corresponding to the perception of Fig. 2.3 had been bistable, Figs. 2.2 and 2.3 would have both featured perceptual switching.

Of these two counterfactuals, [CF_1] appears to be the most plausible. If the differential equations governing our perceptual systems had not featured any bistability, then the dynamics of the neural populations corresponding to those systems would feature none of the see-sawing behaviour predicted by either the oscillator model or the attractor model. That's because if the equations were not bistable, then there would be only one stable perceptual state available for the neural system to enter into. This one stable perceptual state would correspond either to a duck percept, or a rabbit percept. In short, when the differential equations are stripped of the mathematical property of bistability, the corresponding neural system that they represent is constrained with regard to the perceptual states it can realise.

[CF_2] is harder to establish. Suppose the differential equations governing the neural system corresponding to the perception of Fig. 2.3 are bistable. Would it follow that perception switches between two distinct percepts with regard to Fig. 2.3? It is unclear; it really depends on what the bistability of the equations consists in, and whether or not there is some mechanism available to induce switching. If the imagined bistability involves two stable perceptual states, and either the adaptation process or the noise element from the oscillator and attractor models occurs, then switching may well occur. If these conditions are not met, however, then switching is unlikely. In order to work out what the counterfactual situation would be like in this case, we would need to attend more closely to the current neural architecture of human perceptual systems. Doing so would give us some insight into what introducing bistability into a perception of Fig. 2.3 might ultimately do.

Fortunately, there is no need to dig into the details of [CF_2]. The truth of [CF_1] is enough to provide evidence of a difference-making relationship between the mathematical core of the oscillator and attractor models and the explanation for why there is perceptual switching in Fig. 2.2 but not in Fig. 2.3. If [CF_1] is true then it follows that if the mathematical facts had been different, the explanandum – the difference between Figs. 2.2 and 2.3 in terms of switching – would not have been the case. The differential equations in this case therefore help to carry the explanatory load in the explanation.

Indeed, once it is conceded that the mathematical facts are load-carrying with regard to explaining the perceptual difference between Figs. 2.2 and 2.3, it seems to follow that the mathematics is load-carrying in the original switching case for Fig. 2.2. As we have seen, both the oscillator model and the attractor model of the perceptual switching associated with Fig. 2.2 rely on bistable differential equations. Given this, we can formulate a counterfactual that is similar to [CF_1], but that focuses exclusively on the perceptual shifting at issue:

> [CF_3] If the differential equations governing the neural system corresponding to the perception of Fig. 2.2 had not been bistable, perception of Fig. 2.2 would not have switched between a duck percept and a rabbit percept.

[CF_3] appears to be true for the same reason that [CF_1] is. If the differential equations governing a neural system involved in a case of perception of an image like Fig. 2.2 are not bistable, then there will only be one stable perceptual state for the neural system to enter into. If, however, there is only one stable perceptual state for the system to enter into, then only one percept will occur – either duck or rabbit. In short, in a situation in which bistability is removed from the differential equations governing our neural architecture, the perception of an ambiguous figure like Fig. 2.2 turns into something more like the perception of an unambiguous figure like Fig. 2.3. Thus, while the noise and adaptation processes explain the switching itself, the mathematical properties of our neural architecture produce the conditions necessary for switching to occur. This is the explanatory contribution that the mathematical core of those models makes, a contribution that is reflected in the counterfactual structure of the explanation.

At this point, one might raise the following further concern. While it may be right to say that [CF_3] is true, the only way to remove bistability from the differential equations governing the neural system corresponding to the perception of Fig. 2.2 is to make some physical change to the underlying neural system. For it is only by making some physical change to the underlying system that we can properly bring it about that perceptual switching no longer occurs. But whatever that underlying physical change is, surely it is the physical parameter being changed that is ultimately responsible for the perceptual switching at issue. As before, the differential equations are merely being used to ‘index’ or ‘represent’ this underlying physical fact.

In other words, it may be that the apparent alteration to the outcome of the equations specified in the antecedent of the counterfactuals is just an indirect way of changing the physical system that those equations represent. If that’s right, then the counterfactuals themselves underdetermine what the difference-makers are. It may be that the mathematics is playing a merely representational role in the very counterfactuals used to test the explanatory capacity of the mathematics.

What we need, then, is a way to be sure that our counterfactual test is probing the explanatory potential of the mathematics, as opposed to something that the mathematics represents. One way forward is to consider counterfactuals that involve altering the mathematics itself. Thus, we might consider a counterfactual in which we hold fixed that a certain set of differential equations applies to a physical system (such as the equations introduced above that apply to the human perceptual system) and then consider mathematically impossible scenarios in which those equations have less than two roots. If, when considering these scenarios, there is reason to suppose that the switching behaviour that we observe would not occur, then perhaps that is some reason to suppose that the mathematics is making a difference.

The question, though, is whether we can make sense of counterfactuals along these lines and, if we can, how evidential they really are. There is a potential epistemic bootstrapping problem on the horizon. For it may be that we only have reason to suppose that mathematical impossibilities are relevant to working out what makes a difference to the physical world if we already have some reason to believe that mathematics carries the explanatory load in at least some scientific explanations.

There is clearly more to be done, both in terms of gaining a better understanding of counterfactuals that involve mathematics, and in thinking through the evidential relationship between difference-making and counterfactuals. As I said, my goal is not to establish that mathematics really is load-carrying in the bistability explanations offered here. I offer the counterfactual approach simply as an example of the kind of approach one might take to settle the issue.

2.5 Conclusion

It is time to take stock. I have argued that cases of perceptual bistability are important cases of extra-mathematical explanation. In cases of perceptual bistability, a certain perception switches between two stable percepts. I have focused on just one case of bistability: the duck-rabbit illusion. But there are many similar cases available in the literature. The models that I have considered here, and the lessons learned about the explanatory importance of mathematics to those models generalises to these other cases as well.

I have suggested that in models of perceptual bistability, the mathematical facts are explanatorily indispensable and the explanandum is purely physical. Since bistable perceptual phenomena are members of a much larger class of scientific explanations involving bistability, there is a potentially large class of cases that can be used as a basis to defend the enhanced indispensability argument. The question, of course, is whether we can establish that the mathematics in such cases is doing more than representing some underlying physical phenomenon. I have suggested one way to try and answer this question, and have highlighted some challenges with the strategy in question. Ultimately, however, a good deal more work is need to identify a viable epistemology for determining the explanatory capacity of mathematics.

References

Baker, A. (2005). Are there genuine mathematical explanations of physical phenomena? *Mind, 114*, 223–238.

Baker, A. (2009). Mathematical explanation in science. *British Journal for the Philosophy of Science, 60*, 611–633.

Baker, A. (2017). Mathematical spandrels. *Australasian Journal of Philosophy, 95*(4), 779–793.

Bangu, S. I. (2008). Inference to the best explanation and mathematical realism. *Synthese, 160*, 13–20.

Baron, S. (2014). Optimization and mathematical explanation: Doing the Lévy walk. *Synthese, 191*, 459–479.

Colyvan, M. (2002). Mathematics and aesthetic considerations in science. *Mind, 111*, 69–74.

Colyvan, M. (2010). There is no easy road to nominalism. *Mind, 119*, 285–306.

Leng, M. (2005). Mathematical explanation. In C. Cellucci, D. Gillies (Eds.), *Mathematical reasoning, heuristics and the development of mathematics*. London: King's College Publications.

Lyon, A. (2012). Mathematical explanations of empirical facts, and mathematical realism. *Australasian Journal of Philosophy, 90*, 559–578.

Lyon, A., & Colyvan, M. (2008). The explanatory power of phase spaces. *Philosophia Mathematica, 16*, 227–243.

Moreno-Bote, R., Rinzel, J., & Rubin, N. (2007). Noise-induced alternations in an attractor network model of perceptual bistability. *Journal of Neurophysiology, 98*, 1125–1139.

Moreno-Bote, R., Shpiro, A., Rinzel, J., & Rubin, N. (2013). Alternation rate in perceptual bistability is maximal at and symmetric around equi-dominance. *Journal of Vision, 10*. https://doi.org/10.1167/10/11/1.

Peressini, A. F. (1997). Troubles with indispensability: Applying pure mathematics in physical theory. *Philosophia Mathematica, 5*, 210–227.

Saatsi, J. (2016). On the "indispensable explanatory role" of mathematics. *Mind, 125*, 1045–1070.

Shpiro, A., Curtu, R., Rinzel, J., & Rubin, N. (2007a). Dynamical characteristics common to neuronal competition models. *Journal of Neurophysiology, 97*, 462–473.

Shpiro, A., Moreno-Bote, R., Bloomberg, S., Rubin, N., & Rinzel, J. (2007b). Maximum alternation rate in bi-stable perception occurs at equidominance: Experiments and modelling. *BMC Neuroscience, 8*, 78.

Shpiro, A., Moreno-Bote, R., Rubin, N., & Rinzel, J. (2009). Balance between noise and adaptation in competition models of perceptual bistability. *Journal of Computational Neuroscience, 27*, 37–54.

Steiner, M. (1995). The applicabilities of mathematics. *Philosophia Mathematica, 3*, 129–156.

Chapter 3
Description, Explanation and Ontological Committment

Concha Martínez-Vidal and Navia Rivas-de-Castro

Abstract In this chapter, we assess the latest developments of the debate around the "Enhanced Indispensability Argument" by Baker (Philos Math 25:194–209, 2017) and Knowles and Saatsi (Erkenntnis:1–19. https://doi.org/10.1007/s10670-019-00146-x, 2019) to conclude that neither part succeeds in their ambition to overtake the other.

We also aim to clarify what is at stake by looking at the metaphysical pictures that result from the criteria for ontological commitment following from the characterizations of substantial mathematical explanation provided by our authors.

We conclude that (i) the strategy of relying on independently established theories of explanation does not settle the issue. Rather, the choice of one or another theory of explanation, or the way in which the theory is interpreted, seems to be inspired by each author's preferred metaphysical view on mathematical entities; (ii) The explanatory version of the indispensability argument will not take the realist further than its more traditional version since it fails to establish that mathematical entities make a difference. Nevertheless, our nominalists both fail to establish their view and to disallow a moderate version of realism like the one advanced by Baron (Br J Philos Sci 70(3):683–717. https://doi.org/10.1093/bjps/axx062, 2019). There seems to be no fact of the matter to decide between these last two views. Baker 2003, *pace* Philos Math 25:194–209, 2017, seems to be right, in that even if mathematics is indispensable, now *explanatory* indispensable, some version of "the status of No-Difference is unclear, perhaps irredeemably so"

Keywords Enhanced indispensability argument · Non-causal explanation · Topic generality · Cognitive salience

---

This contribution has received financial support from FEDER/Spanish Ministry of Economy and Competitiveness under the project FFI2013-41415-P, and from FEDER/Spanish Ministry of Science, Innovation and Universities - State Research Agency under the project FFI2017-82534-P.

C. Martínez-Vidal · N. Rivas-de-Castro (✉)
University of Santiago de Compostela, Santiago, Spain
e-mail: mconcepcion.martinez@usc.es

J. L. Falguera, C. Martínez-Vidal (eds.), *Abstract Objects*, Synthese Library 422,
https://doi.org/10.1007/978-3-030-38242-1_3

3.1 Introduction

The initial version of the indispensability argument, the so-called Quine-Putnam indispensability argument, aims to establish that we are committed to all and only those mathematical entities that are indispensable for science.[1] The argument has been considered as the best available argument for mathematical (ontological) realism. Nevertheless, there is no consensus about its success. As a result, there is an enormous amount of literature for and against it. In this chapter we discuss the version of the IA proposed by Baker: The Enhanced Indispensability Argument.

Back in 2005, Baker argued that the existence of mathematical objects *does* make a difference in terms of the well-known cicadas' case. The idea underlying this case, as well as other well-known ones such as the Königsberg Bridges, the honeycomb cells, etc., is that these are cases in which mathematics itself plays an explanatory role. Based on these sorts of cases in which mathematics, purportedly, is *explanatorily* indispensable, Baker offered the Enhanced Indispensability Argument (from now on, EIA). This is a version of the indispensability argument that does not include confirmational holism among its premises, and infers the existence of mathematical entities from their *explanatory* indispensability:

> "(1) We ought rationally to believe in the existence of any entity that plays an indispensable *explanatory* role in our best scientific theories.
> (2) Mathematical objects play an indispensable explanatory role in science.
> (3) Hence, we ought rationally to believe in the existence of mathematical objects." (Baker 2009, 613)

The primary idea is that just as scientists infer the existence of a given theoretical entity because it best explains a certain phenomenon—since the existence of the entities 'postulated' in the explanation makes a difference—, we should conclude that mathematical entities exist because they figure indispensably—and hence make a difference—in our best explanations of empirical facts.[2]

Nominalists and platonists have tried to, respectively, block or push the analogy with theoretical entities in scientific explanation by distinguishing levels of mathematical involvement in it.[3] According to Yablo (2012, 1020–22), explanations are such that the *explanandum* E "is explained as arising out of circumstances C by way of a generalization G that links them." In the case of mathematical explanation, he distinguishes three levels of mathematical involvement: descriptive, structural and substantial. Mathematics is descriptively involved in scientific explanation when the context C, the outcome E or the generalization G that connects them cannot be specified without using mathematics. It helps structurally if it is needed "to run the explanation at the appropriate level of generality" (Yablo 2012, 1020), meaning that

[1] See the introduction to this volume for a brief but illuminating account; Colyvan 2001 is a classic source.

[2] As Baron 2016b (ft. 5) claims "the case for realism about mathematics is supposed to 'piggy-back' on the case for scientific realism. The case for scientific realism, however, is made through explanation. So, the case for mathematical realism should proceed via explanation as well."

[3] See Mancosu 2018 for a general account of the debate.

using mathematics allows us to get the appropriate level of generalization[4]; and it is involved substantively "if it provides the covering generalization." (Idem, 1021).

The authors we are discussing—Baker (2017), Baron (2016a, b, 2019), Saatsi (2011), Knowles and Saatsi (2019)—talk about the descriptive and representational uses of mathematics in explanation to refer, respectively, to the first two levels of mathematical involvement distinguished by Yablo. Their terminologies differ in relation to what corresponds to the third level of mathematical involvement; while Baker talks about a certain kind of "explanatory generality" that is specific to mathematics, Knowles and Saatsi (2019) refer to this kind of generality as 'distinctive' mathematical explanation, and Baron discusses 'genuine' mathematical explanation. The shared assumption underlying the debate is that there are cases of 'substantial' mathematical involvement and that this third level of mathematical involvement—substantial involvement—is needed if the argument is to get through, though not all of them agree that it suffices.

Note that the fact that these authors share this assumption is not trivial, but the result of previous debate. Baker (2017) declares that the nominalist has answered the platonist challenge in three different ways in three stages that are previous to the one we are considering. First, the nominalist (i.e. Field 1980) contended that the mathematical apparatus is dispensable altogether. This implies that, in this stage, (some) nominalists acknowledged that the fact that mathematics is involved at the descriptive level could be ontologically committing. In a second stage, nominalists (i.e. Saatsi 2011) argued that mathematics explains through representation since the arguments in which mathematics plays a part are re-phrasable in nominalistic terms. This implies that representing is not ontologically committing since it is possible to do without mathematics. Finally, they have agreed (Knowles and Saatsi 2019) that mathematics plays a substantive role in scientific explanation —that is, they have come to agree to Premise 2—though they reject that this third level of mathematical involvement is ontologically committing— so they reject Premise 1.[5] Hence, at this point, establishing whether there is such a thing as a level of *substantive* mathematical involvement is not at stake among the authors we will be considering; neither is the fact that mathematics is involved in scientific explanation at the other two levels. Rather, the debate has evolved to concentrate on characterizing and demarcating the substantive level of mathematical involvement in explanation, and the ontological commitments, if any, that derive from it. In other words, the situation resulting from platonist and nominalist proposals seems to be—unsurprisingly—

[4]"To give the inevitable example: What is it about a square peg that allows it to slip into a round hole (Putnam 1995, Garfinkel 1981)? The peg's microphysical make-up involves too much unneeded detail; it would still have fit, had it been made of copper. The peg fits if and because the sides are less than $\sqrt{2}$ times as long as the radius of the hole." Yablo 2012, 1020–21.

[5]This strategy is what we know as *easy-road* nominalism. Note that the easy-road is not a single account, but a series of them: Azzouni (1997, 2004, 2012), Balaguer (1996a, b, 1998a, b), Bueno (2012), Leng (2010, 2012), Liggins (2012), Maddy (1995, 1997), Melia (2000, 2002) and Yablo (2002, 2005, 2012), among others. What they all have in common is the rejection of mathematical Platonism with attempts to provide an easier path to nominalism than that of Field's.

that, though some advances have been made, the debate continues now in relation to the notion of "explanatory" (Baker 2017, 1), "distinctive" (Knowles and Saatsi 2019, 1) or "genuine" (Baron 2016a, b, 2019) mathematical explanation and the resulting ontological commitments.

To start with we appraise the dialectical moves made by Baker (2017) and Knowles and Saatsi (2019), who agree that the debate is at an impasse and explicitly formulate their proposals with the aim of overcoming this situation. Their proposals focus on the question of how to determine the commitments of a substantial explanation.[6] The underlying idea is that it can be done by identifying the explanatory virtues that are peculiar to it.

3.2 Understanding *substantial* Mathematical Involvement in Scientific Explanation

The following versions of the famous cicadas' argument will help us elucidate how our authors diverge even if all of them approve that mathematics does not explain by representing and that the mathematical version of the argument, MES (for Mathematical Explanation in Science), is better than the most sophisticated version of its nominalistic counterpart, NES$_*$[7]:

"MES

(4) Having a life-cycle period which minimizes intersection with other (nearby/lower) periods is evolutionarily advantageous. [biological law]
(5) Prime periods minimize intersection (compared to non-prime periods). [number theoretic theorem]
(6) Hence organisms with periodic life cycles are likely to evolve periods that are prime. ['mixed' biological/mathematical law]
(7) Cicadas in ecosystem-type E are limited by biological constraints to periods from 14 to 18 years. [ecological constraint]
(8) Hence cicadas in ecosystem-type E are likely to evolve 17-year periods."

NES$_*$[8]

"(4) Having a life-cycle period which minimizes intersection with other (nearby/lower) periods is evolutionarily advantageous. [biological law]

[6]Marcus noted in his Marcus 2018 the argument "leaves open the question of how one is supposed to determine the commitments of an explanation. EI[A] refers to the theoretical posits postulated by explanations but does not tell us how we are supposed to figure out what an explanation posits." For a similar point see Knowles and Liggins 2015, 3406.

[7]Baker (2017, 2, 5, 6) formulates MES and NES starting with premise number (4) in order to compare these versions of the argument with the versions he calls 'Cicada MES$_{GEN}$' and 'Skid Patch MES$_{GEN}$' (respectively, pages 8–9 and 11).

[8]NES$_*$ provides a general argument pattern: "For any given range [T_1, ..., T_2], of cicada life cycles as determined by the local ecological constraints there is a corresponding instantiation of schema (5/6)$_{**}$. If this instantiation of the argument pattern is true of the given range [. . .], then it can be used to explain the life-cycle length of the given cicada"(Baker 2017, 6).

(5/6)$_{**}$ There is a unique intersection minimizing period T_x for periods in the range [T_1, ..., T_2] years. [fact(?) about time]

(7)$_*$ Cicadas in ecosystem-type E are limited by biological constraints to periods in the range [T_1, ..., T_2] years. [ecological constraint]

(8)$_*$ Hence cicadas in ecosystem-type E are likely to evolve T_x-year periods."

Janssen and Saatsi (2017) analyze the dimensions of mathematical abstraction present in substantive uses of mathematics in explanation. Two of those dimensions have to do with independence: mathematics puts in focus properties or relations that are independent of the "physical constitution" of the objects and from the laws or rules that govern them. The third-dimension abstracts by establishing dependence relations between the objects of study (showing dependencies in the sense of answering what-if questions). Since the three dimensions of abstraction are present in any substantive use of mathematics in explanation, the question lies in determining which dimension plays the relevant explanatory role, how each explanatory virtue relates to these dimensions and their respective ontological implications.

3.2.1 Baker's Account of Substantive Mathematical Involvement in Explanation

Baker (2017) defends that mathematics explains in MES by providing very general information that cannot be understood in terms other than involving mathematical facts and entities. His contention is that this sort of mathematical explanation carries information about worldly features, about the most general features of reality. He distinguishes two kinds of generality: scope and topic generality. The former allows us to (taking the famous cicadas' argument) "express facts about durations and intersection minimization that are unlimited in their scope" (Baker 2017, 5) and that can be used "to collect together, generalize over, and make assertions about, various sets of physical facts in a given domain." (Baker 2017, 7) This corresponds to the first dimension of explanatory abstraction, the one Janssen and Saatsi (2017, 4) characterize as "a matter of relative independence of the *explanans* from the actual physical structure of the entities involved." In Hitchcock and Woodward's (2003, 181) words, it has to do with "generality with respect to objects and systems other than the one that is the focus of explanation". This is the kind of generality emphasized by representational nominalists who contend that scope generality can be guaranteed without using mathematics. Scope generality will not do the job for the realist/platonist/traditional platonist. Baker's challenge to the nominalist comes via 'topic generality'. He pleads (Baker 2017, 7–8) it essential to substantial mathematical involvement in explanation, orthogonal to scope generality and abstracting from the physical constitution of objects—the second dimension of abstraction that Janssen and Saatsi (2017, 4) distinguish (Baker 2017, 8).

To expand on the relevance of this aspect of mathematical explanation and make explicit that mathematical entities make a difference, he gives a new formulation of the cicadas' argument where he moves away from picking out specific facts about time periods and uses the more general notion of 'unit cycles', Cicada MES_{GEN}.[9] Cicada MES_{GEN} brings to light how mathematics goes beyond representation by adding a series of premises that include the germane mathematical facts and how those apply to the cicadas' case. In the same line, he proposes a second, still more general, version of the argument—Skid Patch MES_{GEN}[10]— that applies to rear wheel of brakeless fixed-gear bicycles (BFGBs) and gear ratios. Baker operates this strategy to reveal that the explanations of the cicada case and the BFGBs share the same explanatory core, the first three premises, which state mathematical facts about unit cycles. His conclusion is that topic generality involves non-physical entities and allows unificatory power in a way that is impossible to achieve by the nominalistic versions of the argument.

Baker contends that explanatory depth is also a problem –not so serious, as we shall see—for the nominalist. Given that our nominalist accepts that MES is a better version of the argument, she needs to show how it is that MES, the concept of primeness, explains without entailing any ontological commitments. While NES_*, which does not include any mathematics, "leaves unexplained why this 'fact(?) about time' holds for some ranges of durations and not for others" (Baker 2017, 6), the MES versions of the argument (MES, Cicada Mes_{GEN} or Skid Patch MES_{GEN}) account for it.

In that account, Baker seems to adopt the following criterion for ontological commitment:

> (BOC) We ought to be committed to mathematical objects because they are indispensable for our best explanations, which are those that allow for topic generality, unification, and explanatory depth.

Mathematics substantially explains by abstracting on the properties of physical entities, by unifying different domains and providing explanatory depth (unveiling the deep structural properties of reality). This abstraction or unificatory power is ultimately what justifies our belief in the existence of mathematics since relying on physical entities would make this explanatory task impossible. Therefore, giving that they are indispensable to mathematical explanation –more specifically, these abstraction and unification roles - mathematical entities must exist.

As we shall see below, Knowles and Saatsi offer a genuine answer to his challenge.

[9] Baker 2017, 9–10.

[10] Baker 2017, 11.

3.2.2 Knowles and Saatsi on Substantive Mathematical Involvement in Scientific Explanation

Knowles and Saatsi's strategy involves clarifying the issue in terms of an independently well-established theory of explanation, namely, the counterfactual theory of explanation. They contend that the counterfactual theory of explanation is neutral in that it does not settle in advance whether substantial mathematical explanation is ontologically committing; this sounds, in principle, reasonable since it is an ontic theory of explanation[11] and a monist account, one that intends to provide a unified account for causal and non-causal explanations (Reutlinger 2017, 74).

At the same time, so they go, the fact that the counterfactual theory has an independently gained reputation provides their analysis with further argumentative force and should result in their overcoming the current debate.

According to the counterfactual theory, explanations include *explanandum* and *explanans*. The *explanans* includes generalizations and auxiliary statements (initial conditions, boundary assumptions, theorems, . . .).[12] Moreover (i) the *explanandum*, the generalizations and auxiliary assumptions in the *explanans* have to be (at least, approximately) true; (ii) the *explanans* has to logically entail the *explanandum*[13]; and, (iii) the covering generalization has to establish at least one counterfactual dependency between the auxiliary hypothesis and the *explanandum*. Explanatory counterfactuals capture objective, mind-independent modal (logical) connections. This is the third dimension of abstractness characterized by Janssen and Saatsi (2017), the one that, according to the counterfactual theory, does the explanatory work by showing how the different relevant variables depend on each other, how the *explanandum* would have been different, had the *explanans* been different. In our example, had the cicadas' life-cycle period been different, the duration of the period that minimized intersection would have been different too. The mathematical theorem, in our cicadas' case, the number-theoretical theorem, premise (5) in MES, belongs in the set of auxiliary statements that establish initial or boundary conditions while the mathematics is used also in the covering generalization: premise 6. Given the counterfactual account, both must be true.

Now, the counterfactual account (Woodward 2003) dissects explanatory power in terms of five features of explanations; the relevant ones for mathematical explanation are non-sensitivity, degree of integration and cognitive salience. Two of those three features of explanation, that correspond to Baker's scope and

[11]In their own words: "Ontic accounts take an explanation's explanatory power to at least partly derive from its latching onto worldly things that bear an objective, explanatorily relevant relation to the explanandum." (on line version previous to Knowles and Saatsi 2019)

[12]We follow Reutlinger's description in his Reutlinger 2016 who mentions, "Nagelian Bridge laws, symmetry assumptions, limit theorems, and other modeling assumptions" (Idem, 738) as examples of auxiliary assumptions."

[13]"or a conditional probability $P(E|S_1, \ldots, S_n)$ – where the conditional probability need not be 'high' in contrast to Hempel's covering-law account." (Idem ft.14)

topic generality—respectively non-sensitivity and degree of integration—, and they are about worldly issues; the third one—cognitive salience—is about how the information is conveyed. The counterfactual theory allows mathematics explains by providing information, but "only increases in information about the explanatorily relevant worldly features—objective explanatory dependences—as opposed to changes in how such information is presented, stand a chance of securing further ontological commitments." (Knowles and Saatsi 2019, 18).

The thrust of their paper is to argue, *pace* Baker, that the sort of topic generality that mathematics contributes in substantive explanations fails to convey any ontological commitments because it does not bring to light any worldly features. They do not clarify explicitly what 'wordly' features are, but they agree to the representational nominalist that scope generality can be accounted for in counterfactual terms without including mathematical vocabulary (NES_*). In consequence, they identify 'wordly features' with the (causal) ones depicted in the covering generalizations, those having to do with picturing objective explanatory dependencies.

As for explanations increasing in topic generality, such as the one proposed in Skid Patch $\mathrm{MES}_{\mathrm{GEN}}$, Knowles and Saatsi, once more *pace* Baker, insist that mathematics is not always required to achieve the relevant level of generality. However, they do not demonstrate this point in their paper and for the sake of argument, they are ready to grant that mathematics is sometimes indispensable to express an explanation maximizing topic-generality, to proceed to question the kind of explanatory contribution this kind of theoretical integration would make (Idem, 15).

On one hand, they correctly argue that making an explanation more general "does not help understand the original *explanandum* better" (Idem, 18); rather, the more general the explanation it is not necessarily the better "because it broadens the explanatory landscape". A biologist would not prefer an explanation that integrates different domains just because it is more unified; no doubt, she will be interested in how it applies to *her* domain. Also, making explanations more abstract to increase integration between separate areas sometimes results in a loss of domain-specific detail, and thus loss of explanatory virtues such as factual accuracy or cognitive salience (cf. Ylikoski and Kuorikoski 2010: 213–214). Hence, the fact that a given mathematical result applies in different domains, the fact that mathematics is unificatory in Baker's terms, is of no scientific value, independently of pragmatic or local considerations. They link Baker's emphasis on unification to a return to an implausible Quinean confirmational holism. It seems to us that the fact that mathematics abstracts from different aspects of reality explains why it applies in different areas of knowledge, but it does not spell out 'substantial' mathematical involvement in explanation.

As an answer to Baker's challenge from topic generality, they allege that mathematics adds information resulting in cognitively salient explanations, a virtue that their nominalistic counterparts lack. Still, mathematics does not yield an increase of ontologically committing explanatory information. Their idea is that there are dependencies among physical facts, and mathematical explanations capture them by abstracting on the other two dimensions of nominalistic facts distinguished by Janssen and Saatsi; moreover, mathematics allows to extend the scope of the counterfactuals to any possible duration, and explains by providing information about dependencies among the different variables: prime periods minimize intersection. As a result, mathematics presents information in a cognitively-salient but not-ontologically-committing way.

The idea behind the notion of "cognitive salience"[14] is that human beings as explaining creatures have limited cognitive capacities, and these limitations (partly) determine which explanations are preferable depending on the capacity to enable explainers to draw counterfactual inferences for different values of the *explanans* variables. The verdict is that mathematics' indispensable role in empirical explanations provides no reason to believe in mathematical objects.

Knowles and Saatsi consider that Baker's appeal to explanatory depth involves bringing in the metaphysical picture. But, given the naturalistic roots of the EIA debate, there is no reason to take an appeal to explanatory depth seriously (Knowles and Saatsi 2019, 15). There is no hard evidence that scientists consider mathematical explanations better because they are sensitive to what platonists see as mathematics' contribution to these explanations (cf. Knowles and Liggins 2016: 3404). Hence, the appeal to scientific practice and the beliefs of scientists cannot be used as an argument by platonists for such a metaphysical commitment. They also contend that the value of explanatory depth cannot be demonstrated. The degree to which a subject finds explanatory depth metaphysically illuminating will depend on her background metaphysics, so, for instance, a Humean-spirited person will not find this kind of depth explanatory.

Thus, Knowles and Saatsi articulate an account for mathematical explanation of empirical phenomena while the claim that many explanations involving mathematics are more powerful than their nominalistic counterparts remains unchallenged.

It follows that they offer the following criterion of ontological commitment:

> (KSOC) We ought to be committed to those objects that feature in objective (wordly) explanatory dependencies.

According to it, the EIA fails to establish the existence of mathematical entities.

Nevertheless, their dialectical move fails, since the counterfactual theory demands that the covering generalizations and auxiliary statements be true; but they are contending that they provide cognitive salience, hence they are not (approximately) true, and they are not ontologically committing. As a result, it is not clear that the counterfactual theory provides an independently established theory of explanation or that it is neutral about the role of mathematics in explanation. If, as Reutlinger (2017, 79–80) suggests, non-causal explanations have to do with *understanding* how things could have been, then, the counterfactual theory ceases to provide a neutral ground. Moreover, Reutlinger himself acknowledges that the counterfactual theory of explanation needs to be modified to account for non-causal explanation.

[14] See Ylikoski and Kuorikoski 2010 p. 215.

Additionally, even if the counterfactual theory succeeded, Knowles and Saatsi acknowledge they need to show that the counterfactual theory offers the only way of understanding the role of mathematics in substantive explanation; but, in what follows we argue that Baron's proposal providesan provides an alternative view. Moreover, he contends that the counterfactual theory is not adequate to account for 'genuine' (Baron's term for what we have been calling 'substantial') mathematical explanation.

3.2.3 Baron on Genuine Mathematical Explanation

Baron is interested in those explanations in which "a non-mathematical fact is explained… by mathematical facts" (Baron 2019, 683); and, agrees that what is essential to the substantial role of mathematics in explanations is that it plays a non-descriptive role (Baron 2016a, b, 2019).

Two aspects of Baron's proposal are relevant to our argument: (1) he rejects that the counterfactual theory of explanation can account for non-descriptive uses of mathematics in explanation; (2) his account of genuine mathematical explanation relies on a different independently established theory of explanation that illustrates how mathematics contributes by either allowing for a certain kind moderate platonism (thus conveying that the sort of platonism defended by Baker, —as we shall argue below—heavy-duty platonism, does not obtain) or for a certain kind of nominalism (thus establishing that nominalism is one of two possibilities).

He illustrates non-descriptive uses of mathematics in explanation with some well-known cases in the literature (the cicadas', Könisberg Bridges, etc.) and contrasts those with descriptive uses of mathematics. To illustrate the latter, he offers the case in which we use mathematics to determine the time of arrival of a train, given the departure station, the speed of the train, and the distance to the arrival station.[15] As we have already mentioned, his view is that the counterfactual account does not allow for the distinction.[16]

The contrast between both analyses in terms of the counterfactual theory is not obvious since the counterfactuals both parties are considering are dissimilar. Contrary to Knowles and Saatsi, who analyse counterfactual dependencies between empirical facts, Baron formulates counterfactual dependencies between mathemati-

[15]The train case goes as follows: "… suppose we want to explain why a train T arrives at a station, S, at 3:00 pm. The explanation is as follows: T left another station, S_*, 10 kilometers away at 2:00 pm and headed towards S at 10Kph. Obviously, this explanation exploits some basic mathematics. Numbers are used to state the distance between stations as well as the speed of the train and a very basic mathematical calculation is deployed, namely $10/10 = 1$. However, the mathematics itself does not do any explanatory work. Rather, it is facts about the speed of the train and the distance between stations that fully explain why the train arrived when it did. The mathematics just helps us to express the explanatory facts at issue." (Baron 2016a, b, 460)

[16]See Baron 2016a, b, 468–473.

cal and non-mathematical facts: had the mathematics been different, physical things would have been different.[17] His view is (Baron 2016a, b, 468–474) that the sort of dependencies between mathematical and physical facts that obtain are the same in descriptive and non-descriptive uses of mathematics, the train and the cicadas' case; hence, the counterfactual account is not appropriate to articulate the peculiarity of non-descriptive uses in an ontologically committing way. Baron (2016a, b, 474) himself acknowledges this result might show that substantial uses of mathematics in explanations should not be accounted for in terms of their providing non-descriptive information that is ontologically committing. Thus, Knowles and Saatsi could contend that Baron's failure shows that the counterfactual theory of explanation applies in their fashion. But we have argued above that the counterfactual theory of explanation in the way they understand it, does not directly apply to the non-causal case.

Baron (2019) provides a characterization of the distinction between descriptive and non-descriptive uses of mathematics in terms of his mathematical version of the Deductive-Nomological theory of explanation. According to the Deductive-Nomological theory,[18] an explanation is an argument whose premises are the *explanans* and whose conclusion is the *explanandum*; the *explanans* comprises a set of initial conditions and a set of laws of nature. Baron introduces some details in order to adjust it to the case of genuine mathematical explanation. First, genuine or extra-mathematical explanations are sound relevant-arguments (R-arguments),[19] "where an R-Argument is an argument in which all of the information contained within the conclusion of the argument is contained within the premises, and each premise contributes some part of the information contained within the conclusion." (Baron 2019, §8) As is to be expected, the conclusion of an extra-mathematical explanation is a proposition stating the physical phenomenon to be explained. The premises of an extra-mathematical explanation should include, apart from those describing the initial conditions, at least one mathematical claim. This condition has to do with the fact that he is adapting the Deductive-Nomonological account to non-causal explanation.

This mathematical claim must be such that, because of including it, the obtaining extra-mathematical explanation satisfies the genuineness condition. In addition, to be genuine, it must comply with "The Razor-Sharp Essential Deducibility

[17] See Baron 2016a, b, 468).

[18] See Woodward and Hitchcock 2003.

[19] He chooses relevant logic as background logic. The reason for that is that classical logic is monotonic, and since mathematical theorems are true in every model, then, it is possible to add any mathematical theorem to our set of premises and we still get a sound explanation, but one in which irrelevant information can appear. Anyway, he acknowledges that it is possible to adopt classical logic as a background logic by adding the following proviso:

> Containment: The premises of an extra-mathematical explanation must contain all of the information contained within the conclusion and each premise must contribute some part of that information. (Baron 2019, 704)

Constraint (REDC)"[20] and pass the informational test. These constraints are meant to guarantee that they do not collapse in descriptive uses of mathematics. If they did, he claims, and it is obvious, the EIA would collapse in the Quinean version of the indispensability argument.

The Razor-Sharp Essential Deducibility Constraint (REDC) demands that explanations that include among its premises mathematical information are better than any possible nominalistic correspondent is. Baron characterizes better explanations in terms of explanatory virtues that are analogous to those put forward by Baker and Knowles and Saatsi: simplicity and unity. The best arguments are those in which an optimal balance between simplicity and unity obtains.[21] Simplicity has to do with the number of premises of the argument, where the smaller the number of premises, the better the argument. Given two arguments A and A_* with the same conclusion C, A is better than A_* if A's premises can be used to establish a greater range of conclusions other than C.[22,23]

The informational test (Baron 2019, 703–4), which we simplify here, contributes to ensuring that the mathematical proposition includes non-descriptive information: If we remove the part of the information I that is descriptive about a physical system while holding the rest of the informational content of M fixed, two things can happen. If (i) the information in I exhausts the information that M carries regarding the *explanandum*, then M fails the test and is only contributing descriptive information to the explanation. If, on the other hand, (ii) the information in I does not exhaust the information that M carries regarding the *explanandum*, then M passes the test and carries non-descriptive information regarding the *explanandum*, that is, information that is used to derive the *explanandum* and thus explain it. In the first case, M is playing a merely descriptive role in the explanation. In the second case, M plays a non-descriptive role in the explanation. If, furthermore, there are no alternative derivations of the *explanandum*, M is indispensable to explaining it. In the case where M fails the informational test, even if it is indispensable to the explanation, it is only contributing descriptive information. If it passes the

[20] "A non-mathematical claim P is essentially deducible from a premise set S that includes at least one mathematical sentence M just when for an appropriate choice of expressive resources there is a sound derivation of P from S and either for the same choice of expressive resources there is no sound derivation of P from a premise set S_* that includes only physical sentences or all sound derivations of P from premise sets $S_1 \ldots S_n$ each of which includes only physical sentences are worse than the mathematical derivation or for all appropriate choices of expressive resources the best derivations use M." (Baron 2019)

[21] Following Lewis (1983), he also refers to 'unity' as strength.

[22] See Baron 2019 §7 for the problems his proposal faces in relation to unity and how they are to be solved.

[23] It is important to note, that Baron thinks that different explanations are to be compared in terms of the consequences they have for physical claims only. This last restriction has to do with the fact that he intends to characterize extra-mathematical explanation, he is not interested in mathematical explanations of mathematical facts.

informational test, then M is both indispensable to the explanation and playing a genuine explanatory role in the explanation.

The account provides an elucidation of how the *explanans* carries non-descriptive information regarding the *explanandum*, and this information is used to derive the *explanandum* and thus explain it. In this case, the *explanans* plays a non-descriptive role in the explanation.

Baron introduces two possible ways of making sense of how mathematical claims contain information about physical systems (Baron 2019, §§7–8): (i) mathematical claims carry physical information "indirectly, via structural mappings;" (ii) or physical claims carry mathematical information "in virtue of physical objects possessing mathematical properties."

He clarifies the first in terms of Bueno and Colyvan's inferential conception of applied mathematics (2011, 353–354). They distinguish three stages in the application of mathematics: (i) immersion: whereby a physical system is mapped into a mathematical structure, (mathematics represents); (ii) inference, whereby inferences are drawn within mathematics about the relevant mathematical structure (results in mathematics are considered); and, (iii) interpretation, whereby the results of the inferential step are interpreted physically. This does not convey that there is a mathematical-free description of the empirical domain.[24] But this provides a sensible account of how mathematics explains, one that does not rely in its unificatory power—on the contrary, it assumes it—and is consistent with the idea that a mathematical explanation works for a scientist in as much as it clarifies what is going on in her domain, not by the fact that a given mathematical theorem applies in other fields as Baker contends.

Hence, although mathematics is used to describe a physical system via a structural mapping, the mathematics does more than merely describe (2017, 32) because claims about mathematical structures are deduced and used to provide non-descriptive information about physical systems via the interpretation step. In our cicadas' case, the number-theoretic theorem provides information about why 17-years periods minimize intersection.

Just as laws of nature carry information about idealized systems and manage to provide information about particular systems only when a certain structural mapping relation is established, "[m]athematical claims carry information in much the same manner" (Baron 2019, 707). They encode high-level structural information by carrying information about some abstract object (a structure) and then via a structural mapping to physical objects. (Idem) The mathematical structure contains information about the physical, about worldly features.

[24] "The empirical setup is the relevant bits of the empirical world, not a mathematics-free description of it." (Bueno and Colyvan 2011, 354) "The results are read back down into the physical system via some structural mapping relation between the mathematical and physical structures. The structural mapping used in the interpretation step need not be the inversion of the mapping used at the immersion step." Baron (2019, 33) refers to Batterman 2010 for an account of cases in which no immersion is possible.

As a result, the ensuing criterion for ontological commitment would be something like:

($BaOC_1$) We ought to be committed to mathematical facts because they carry non-descriptive information about wordly features.

Alternatively, it can be said that "claims about physical systems (sometimes) carry mathematical information. Here is the basic idea: it is not just mathematical objects such as numbers, sets, functions, and so on that possess mathematical properties. Rather, some physical objects also possess mathematical properties. [. . .] First example: Space-time is thought to be a physical object. However, it is also thought that space-time is continuous. Continuity, however, is a mathematical property. Therefore, spacetimes have mathematical properties. Second example: Physical objects possess shapes. The property of having a certain shape, however, is a mathematical property, enshrined in geometry and describable algebraically. So, physical objects have mathematical properties." (Baron 2019, 710–11).[25]

The ensuing criterion for ontological commitment would be something like:

($BaOC_2$) We ought to be committed to mathematics because physical entities have mathematical properties.

Hence, $BaOC_1$ seems to have contributed an account of how mathematics explains that is ontologically committing. Still, the metaphysical status of the ensuing entities needs to be clarified.

For instance, his second criterion, $BaOC_2$, seems to convey that (in Liggins 2016 words) "the obtaining of quantitative relations is explained by the objects' possession of nominalistic properties."

3.3 From Their Criteria for Ontological Commitment to the Underlying Metaphysical Pictures

The ontological commitments of a sentence are the ontological demands on the world that derive from the truth of the sentence; those cannot be changed without changing the truth-conditions of the sentence.[26] Subscribing to premise one in the EIA conveys believing in the existence of those entities that play an indispensable explanatory role in MES. All our authors subscribe to the entities that are explanatorily indispensable according to their criteria. Likewise, implicit to our previous discussion is that they concur about the ontological commitments

[25] He claims that this view "dovetails nicely with Rizza's (2013) account of applied mathematics." Rizza's account, as described by Baron, contends that some cases of applied mathematics operate by first identifying a formal property of a physical system; second reasoning mathematically about the physical system; and third using those mathematical results to get further detail about the particular physical system.

[26] We follow Rayo 2007.

stemming from the different explanatory virtues and differ about their relevance to substantial mathematical involvement. It remains for us to explore the metaphysical pictures underlying their criteria for ontological commitment.

3.3.1 Baker's Criterion for Ontological Commitment

Baker's interest in those versions of the argument that make MES' commitment to mathematical entities explicit are meant to unveil the real commitments of those arguments in which mathematics plays a substantial role. Given that only those versions of the cicadas' argument in which an explicit mention of non-physical entities (unit cycles) satisfy his criterion, it follows that mathematical entities exist. Clearly, he intends to show that the existence of mathematical entities makes a difference to the concrete world. But how can mathematical objects make a difference to physical objects? What is the underlying metaphysical picture?

Plebani (2018) contends that the obtaining view at the metaphysical level determines how to conceive of abstract objects. In particular, he argues that the EIA needs something like heavy-duty platonism. Heavy-duty platonism considers that how concrete—physical—entities are fundamentally depends on how mathematical entities are.[27] Hence, the mathematical and the physical cannot be orthogonal, that is, they cannot be metaphysically independent.[28]

Liggins (2016, 534) reassesses the EIA in terms of the grounding relation and discusses three different ways of conceiving of relations between quantitative (descriptive or representational) and nominalistic facts. We will consider the two relevant ones: (i) "there is no explanation", "the relation is a primitively metaphysical one; and, (ii) "the obtaining of quantitative relations is explained by the objects' possession of nominalistic properties." As we shall see, only the former allows for mathematics making a difference.

Liggins claims that for unification to be really explanatory, relations between mathematical and empirical facts need to be metaphysically fundamental, meaning that "the relation does not obtain in virtue of anything" (Liggins 2016, 534) or, in Knowles words, the relation is "not derivative of [. . .] properties or relations that hold of physical objects alone." (Knowles 2015, 1255). Indeed, Baker's point is precisely that mathematics explains in a unificatory way because how concrete things are depends on how mathematical facts are. This seems to be the idea underlying Baron's way of applying the counterfactual theory of explanation.

[27] Knowles and Liggins 2015; Liggins 2016.

[28] Moderate versions of platonism related to the original indispensability argument, for instance the one advocated by Hellman and Putnam, demand that the mathematical and the concrete are orthogonal, therefore, in that case, the mathematical does not make a difference. See Martínez-Vidal 2018.

Liggins rejects heavy-duty platonism by posing two problems to it. First, the idea that the way physical objects are depends on how mathematical objects are goes against our pre-theoretical intuitions: we usually think that the way physical objects are does not depend on how mathematical objects are. Second, since there are various ways of measuring weight or distance, there are no good reasons to choose one scale or another as the one that determines how physical things are. Liggins concludes that the Platonist cannot but account for the relation between the mathematical and the physical in terms of the grounding relation which, as we shall see, offers an alternative non-heavy-duty view at the metaphysical level. Nevertheless, Liggins conclusion is far from unanimous. Baker (2003) rejects the argument mentioned by Liggins on the basis that it relies on an analogy with concrete objects that fails: it starts with the a-causal character of abstracta to demand that the way in which abstracta could be relevant to the physical world is by acting in a causal way. Moreover, Knowles (2015) argues that none of the arguments against heavy-duty platonism succeeds. Hence, the possibility that the relation between mathematical and concrete facts is a primitive relation of metaphysical dependence is still alive. A different issue is whether the EIA –Baker's proposal—succeeds; we have argued it does not.

3.3.2 Knowles and Saatsi and Ontological Commitment

Knowles and Saatsi rely on a well-established theory of explanation to conclude that substantial mathematical explanation conveys no ontological commitments to mathematical entities; rather, substantial mathematical explanation provides cognitive salience.

But, as mentioned above, according to the counterfactual theory, an explanation must satisfy several constraints; among them that the *explanans* must be true or approximately true. Hence, they manage to avoid committing to mathematical entities at the price of having to renounce their appeal to the 'independently' well-established theory of counterfactual explanation. The counterfactual theory simultaneously allows them to account for the role of non-descriptive uses of mathematics in epistemic terms, but at the same time it demands the truth of all the premises in the argument and the argument must be the best one, MES.

To overcome this situation they might argue that the counterfactual theory needs to be re-formulated since the demand that the *explanans* be true would be too strong for "how-possibly explanations" as Reutlinger acknowledges (Reutlinger 2017, 79–80); or claim that they are not fully asserting all the premises, given that a paraphrase, NES_*, in our case, is available. It seems to us that their acknowledging that they need to provide formulations for the different cases with enough scope generality (degree of integration) and no mathematics in them in every case is not enough to address this problem. At least, it does not amount to reducing the problem to an easier one.

We have already contended that giving up on the counterfactual theory of explanation has serious consequences for their dialectical move. Rather than relying on an independently well-established theory of explanation that helped push the debate passed the impasse, they have showed that the counterfactual theory has to be modified in order to improve its account of non-causal explanation, as Reutlinger acknowledges.

As a result, it is not clear that they succeed in showing that the existence of mathematical entities makes no difference at the ontological level.

Moreover, since they claim that mathematics is cognitively salient because it abstracts on the properties of objects and of the relations that hold among them, then they acknowledge it provides information about the physical. But this brings us to Baron's account of non-descriptive mathematical information, and we still need to see how this is possible.

3.3.3 Baron and Ontological Commitment

To find out whether Baron's characterization of substantial mathematical involvement in explanation suffices to establish that mathematical objects make a difference, we need to explore the metaphysical pictures underlying the two criteria.

The first one emphasizes that mathematical structures make a difference:

(BaOC$_1$) We ought to be committed to mathematical facts because they carry non-descriptive information about wordly features.

Since he relies on the inferential approach (Bueno and Colyvan 2011) and subscribes to the following version of the indispensability argument,

"(A1) We ought to believe in the truth of any claim that plays an indispensable explanatory role in our best scientific theories.
(A2) Some mathematical claims play an indispensable explanatory role in science.
Therefore, (A3) We ought to believe in the truth of some mathematical claims." (Baron 2016a, b, 458)

he is compromising to mathematical structures. Nevertheless, this does not imply there is some sort of fundamental metaphysical dependence of the kind advocated by heavy-duty platonists.[29] Is it the case that had mathematical structures been different, the facts in the world would have been different (hence, heavy-duty platonism follows)? Or, is it the case that had physical facts been different, the mathematics used to describe them and to make inferences about them would have been different (hence heavy-duty platonism does not obtain)?

[29] He characterizes heavy-duty platonism as stated above: Baron 2016a, b, 460.

Take two of the premises in MES (Baker 2017):

(5) Prime periods minimize intersection (compared to non-prime periods). [number theoretic theorem]
(6) Hence organisms with periodic life cycles are likely to evolve periods that are prime. [‘mixed’ biological/mathematical law]

What mathematics allows us to infer in a transparent and quick way is that certain life-cycle lengths are optimal solutions to improve survival rates. Does it follow from the truth of this assertions that this is so because of a metaphysical fundamental relation? Baron explores this possibility when he tries to account for genuine uses of mathematics in scientific explanation in terms of the counterfactual theory. His conclusion is that the counterfactual theory does not allow us to explain the difference between descriptive and non-descriptive (or informative and non-informative) uses of mathematics in explanation as we discussed above. In addition, Bueno and Colyvan contend that the inferential theory of mathematical explanation is neutral about the ensuing ontological commitments. Hence allowing for non-descriptive uses of mathematics is not enough to establish heavy-duty platonism. Moreover, Baron’s discussion allows for the second criterion formulated above, and that criterion is consistent with a metaphysical view that is clearly more palatable to the naturalist.

The second criterion, ($BaOC_2$) sounds weaker[30] and compatible with moderate Platonism, a position he characterizes using Field’s (1989, 186–93) words: “relations between physical things and numbers are conventional relations that are derivative from more basic relations that hold among physical things alone.”

Liggins (2016) argues for understanding the quantitative relation between mathematical and concrete facts as follows: “the obtaining of quantitative relations is explained by the objects’ possession of nominalistic properties.” In that case, the concrete grounds the mathematical. He illustrates what being grounded means with several examples:

> Socrates’ singleton set exists because Socrates does. The proposition that dogs bark or pigs fly is true because dogs bark. [. . .] The existence of Socrates’ singleton depends on the existence of Socrates, the truth of the proposition depends on the behavior of animals, [. . .] the dependence seems not to be causal in nature. Let us call non-causal dependence, grounding. (Liggins 2016, 532)

In general, relation ‘S’ between the concrete (physical) ‘c’ and the mathematical ‘m’ obtains because the concrete objects have a certain nominalistic property S-m: cSm because S-mc; the mass of this melon in kg is 2,2 because the-mass-in-Kg property holds of this melon. This gives an advantage to the nominalist because she can say that the mathematics represents the concrete. However, substantial mathematical involvement is about non-descriptive uses of mathematics. Hence, we need to see whether concrete facts can ground them.

[30] Idem.

Plebani (2016, 551) explores this possibility, to conclude that concrete facts cannot ground cases such our cicadas' case explained in terms of MES. Plebani refers to the fact that mathematical explanations are better because they are more illuminating. He advances a test to establish whether there are nominalistic facts that ground the obtaining of the relation between the concrete fact and the non-descriptive mathematical fact. The query is which is the general fact that grounds the illuminating explanation according to which prime periods minimize intersection. Plebani explores the possibility that a mathematical fact at an appropriate level of generality could be grounded by a disjunctive fact. Consider the cicadas' case: the being prime of the life-cycle could be replaced by its being 3 or being 5 or being 7, ... Now, since it is commonly accepted for grounding relations that the truth of a disjunction is grounded in the truth of one (at least one) of its members, then the truth of the general fact can be grounded in the truth of one of the members of the disjunction. Hence, the general fact would be grounded in the disjunctive fact, and the disjunctive fact in the truth of one of its disjuncts. Because grounding is a transitive relation, the general fact would be grounded in the primeness of one of the numbers featuring in the disjunction. The fact that prime periods minimize intersection would be grounded in the truth of one of the concrete facts about the duration in years of the cicadas' life cycle in the disjunction. The problem is that the concrete content of the sentence "Organisms with periodic life-cycles are likely to evolve periods that are prime" would be different from what is usually taken to be the concrete content of a sentence, in symbols 'IISII.' Yablo characterizes it in this way:

> IISII is the proposition true in a world w iff S is true in some v concretely indiscernible from w, albeit perhaps richer than w in mathematical objects. (Yablo 2010, 6)

Plebani uses 'ISI' to refer to the full content (concrete and non-concrete) of sentence S. Then, he applies the definition of grounding above:

1. the concrete content of the sentence "Organisms with periodic life-cycles are likely to evolve periods that are prime" would be either this life-cycle has length 3 or this life-cycle has length 5 or this life-cycle has length 7 or ...
2. Then, the ground for the disjunction could be any of the members of the disjunction; for instance, this life-cycle has length 7.
3. Hence, what grounds IISII is different from IISII, and what grounds ISI is different from IISII.

He concludes that what grounds S in this world is not what is conveyed when we assert S. But, at least in our case, this should come as no surprise; after all, paraphrasing Bliss and Trogdon 2016, to say that "[O]rganisms with periodic life-cycles are likely to evolve periods that are prime" is not just to say that organisms with periodic life-cycles are likely to evolve periods that are 17, for periodic life-cycles can have other durations. What is claimed is that the fact that the life-cycle is 17 explains (grounds) why the life-cycle is prime. When we identify what grounds the truth of a sentence at this world, we are not at all pursuing the 'content' (concrete or not) of the sentence. Content has to do with different ways in which a sentence can be true (grounded) at different worlds, worlds in which other members of the grounding disjunction could be the ones explaining why the life-cycle is prime. In consequence, the grounding account might still be alive.

Anyway, what matters to our discussion is that Baron proposes an analysis of substantive mathematical involvement in explanation that is compatible with moderate platonism- though it fails to establish it, since it is also compatible with nominalism. Hence, the view that non-descriptive uses of mathematics in science allow for information to flow from the *explanans* to the *explanandum* does not by itself settle metaphysical issues.

3.4 Conclusion

We have argued that current debate about the EIA fails to be conclusive. Baker's claim in 2003 that "if mathematics is indispensable, then the status of No-Difference is unclear, perhaps irredeemably so" continues to be the case even if we go beyond descriptive indispensability to substantial mathematical involvement.

It is also the case that Baker's challenge to the nominalist in terms of topic generality fails. Topic generality is not the explanatory virtue that accounts for substantive mathematical explanation. Knowing that being prime explains several phenomena in the physical world, does not allow the biologist to explain why cicadas reproduce in prime life-cycles.

Knowles and Saatsi also fail to establish that the mathematical information in substantial mathematical involvement is not ontologically committing but just cognitively salient by appealing to a well-established theory of explanation. Apart from the difficulties pointed out above, the problem, as we have just seen, is that Baron's proposal provides an alternative account of substantive mathematical involvement, also in terms of a well-established theory of explanation that can be interpreted as establishing that certain relations between mathematical and physical facts obtain, but that is inconclusive about the underlying metaphysical picture.

At this point, it should be clear that the way in which the different authors conceive of substantive mathematical involvement in scientific explanation is not independent of their preferred metaphysical view about mathematical objects; it is not independent of their position in front of the EIA. It is also clear that the choice of one or another theory of explanation is guided by that position. What one's preferred theory of explanation is, or what one's explanatory virtue is, is not independent of the sort of account of non-causal explanation one is ready to accept.

Relying on an independent well-established theory of explanation does not allow us to overcome the impasse.

Moreover, it seems that heavy-duty platonism is radically apart not only from nominalism but also from those forms of platonism that are moderate. Neither the latter nor the nominalist allow for any primitive relation of metaphysical dependence between the mathematical and the physical to obtain, while heavy-duty platonism contends that such a relation obtains. In fact, many have come to consider that there is no fact of the matter about the existence or not of mathematical entities (this issue is discussed by Leng in this volume) if ‘existence’ is understood in some sort of ‘moderate’ platonist way. In other words, at this stage of the debate, there seems to be no way to decide between ‘moderate’ platonism of one sort or another and nominalism. Some authors make the point by noting that what is relevant to mathematical practice is not whether mathematical entities exist, but mathematical content.

Acknowledgments We would like to thank an anonymous referee for her very helpful comments and the audience of the Workshop “Abstract Objects?” celebrated in Santiago de Compostela in 2018 for their patience and feed-back.

References

Azzouni, J. (1997). Applied mathematics, existential commitment and the Quine-Putnam indispensability thesis. *Philosophia Mathematica, 5*(3), 193–209.

Azzouni, J. (2004). *Deflating existential consequence: A case for nominalism*. Oxford: Oxford University Press.

Azzouni, J. (2012). Taking the easy road out of Dodge. *Mind, 121*(484), 951–965.

Baker, A. (2003). Does the existence of mathematical objects make a difference? *Australasian Journal of Philosophy, 81*(2), 246–264. https://doi.org/10.1080/713659635.

Baker, A. (2005). Are there genuine mathematical explanations of physical phenomena? *Mind, 114*(454), 223–238.

Baker, A. (2009). Mathematical explanation in science. *British Journal for the Philosophy of Science, 60*, 611–633.

Baker, A. (2017). Mathematics and explanatory generality. *Philosophia Mathematica, 25*(2), 194–209.

Balaguer, M. (1996a). A fictionalist account of the indispensable applications of mathematics. *Philosophical Studies, 83*, 291–314.

Balaguer, M. (1996b). Towards a nominalization of quantum mechanics. *Mind, 105*, 209–226.

Balaguer, M. (1998a). *Platonism and anti-platonism in mathematics*. Oxford: Oxford University Press.

Balaguer, M. (1998b). Attitudes without propositions. *Philosophy and Phenomenological Research, 58*, 805–826.

Baron, S. (2016a). Explaining mathematical explanation. *The Philosophical Quarterly, 66*(264), 458–480. https://doi.org/10.1093/pq/pqv123.

Baron. (2016b). The explanatory dispensability of idealizations. *Synthese, 193*(2), 365–386. https://doi.org/10.1007/s11229-014-0517-z.

Baron, S. (2019). Mathematical explanation by law. *British Journal for the Philosophy of Science, 70*(3), 683–717. https://doi.org/10.1093/bjps/axx062.

Batterman, R. W. (2010). On the explanatory role of mathematics in empirical science. *British Journal for the Philosophy of Science, 61*, 1–25.

Bliss, R., & Trogdon, K. (2016). Metaphysical grounding. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2016 Edition), URL = https://plato.stanford.edu/archives/win2016/entries/grounding/

Bueno, O. (2012). An easy road to nominalism. *Mind, 121*, 967–982.

Bueno, O., & Colyvan, M. (2011). An inferential conception of the application of mathematics. *Nous, 45*(2), 345–374. https://doi.org/10.1111/j.1468-0068.2010.00772.x.

Colyvan, M. (2001). *The indispensability of mathematics*. New York: Oxford University Press.

Field, H. (1980). *Science without numbers*. Oxford: OUP. (2nd edition 2016).

Field, H. (1989). *Realism, mathematics and modality*. UK: Basil Blackwell.
Garfinkel, A. (1981). *Forms of explanation: Rethinking the questions in social theory*. New Haven: Yale University Press.
Hitchcock, C., & Woodward, J. (2003). Explanatory generalizations, part II: Plumbing explanatory depth. *Nous, 37*(2), 181–199.
Janssen, L., & Saatsi, J. (2017). Explanatory abstractions. *British Journal for the Philosophy of Science, 0*, 1–27.
Knowles, R. (2015). Heavy duty platonism. *Erkenntnis, 80*(6), 1255–1270.
Knowles, R., & Liggins, D. (2015). Good weasel hunting. *Synthese, 192*, 3397–3412.
Knowles, R., & Saatsi, J. (2019). Mathematics and explanatory generality: Nothing but cognitive salience. *Erkenntnis*, 1–19. https://doi.org/10.1007/s10670-019-00146-x.
Leng, M. (2010). *Mathematics and reality*. Oxford: Oxford University Press.
Leng, M. (2012). Taking it easy: A response to Colyvan. *Mind, 121*(484), 983–995.
Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy, 61*, 343–377.
Liggins, D. (2012). Weaseling and the content of science. *Mind, 121*(484), 997–1005.
Liggins, D. (2016). Grounding and the indispensability argument. *Synthese, 193*, 531–548.
Maddy, P. (1995). Naturalism and ontology. *Philosophia Mathematica, 3*, 248–270.
Maddy, P. (1997). *Naturalism in mathematics*. Oxford: Clarendon Press.
Mancosu, P. (2018). Explanation in mathematics. IN: E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2018 Edition), URL = https://plato.stanford.edu/archives/sum2018/entries/mathematics-explanation/
Marcus, R. (2018). The indispensability argument in the philosophy of mathematics, In: The internet encyclopedia of philosophy. https://www.iep.utm.edu/indimath/
Martínez-Vidal, C. (2018). Putnam and contemporary fictionalism. *THEORIA: An International Journal for Theory, History and Foundations of Science, 33*(2), 165–181. https://doi.org/10.1387/theoria.17610.
Melia, J. (2000). Weaseling away the indispensability argument. *Mind, 109*, 455–479.
Melia, J. (2002). Response to Colyvan. *Mind, 111*, 75–79.
Plebani, M. (2016). Nominalistic content, grounding, and covering generalizations: Reply to grounding and the indispensability argument. *Synthese, 193*(2), 549–558.
Plebani, M. (2018). The indispensability argument and the nature of mathematical objects. *THEORIA: An International Journal for Theory, History and Foundations of Science, 33*(2), 249–263. https://doi.org/10.1387/theoria.17613.
Putnam, H. (1995). Philosophy and our mental life. In P. Moser (Ed.), *Contemporary materialism: A reader* (pp. 122–132). New York: Routledge.
Rayo, A. (2007). Ontological Commitment. *Philosophy Compass, 2*(3), 428–444. https://doi.org/10.1111/j.1747-9991.2007.00080.x.
Reutlinger, A. (2016). Is there a monist theory of causal and non-causal explanations? The counterfactual theory of scientific explanation. *Philosophy of Science, 83*(5), 733–745. https://doi.org/10.1086/687859.
Reutlinger, A. (2017). Extending the counterfactual theory of explanation. In A. Reutlinger & J. Saatsi (Eds.), *Explanation beyond causation: Philosophical perspectives on non-causal explanations* (pp. 74–95). Oxford: OUP.
Rizza, D. (2013). The applicability of mathematics: Beyond mapping accounts. *Philosophy of Science, 80*(3), 398–412. https://doi.org/10.1086/670295.
Saatsi, J. (2011). The enhanced indispensability argument: Representational versus explanatory role of mathematics in science. *The British Journal for the Philosophy of Science, 62*(1), 143–154.
Woodward, J. (2003). Experimentation, causal inference, and instrumental realism. In H. Radder (Ed.), *The philosophy of scientific experimentation* (pp. 87–118). Pittsburgh: University of Pittsburgh Press.
Woodward, J., & Hitchock, C. (2003). Explanatory generalizations, part 1: A counterfactual account. *Noûs, 37*, 1–24.

Yablo, S. (2002). Go figure: A path through Fictionalism. *Midwest Studies in Philosophy, 25*, 72–102.

Yablo, S. (2005). The myth of the seven. In M. Kalderon (Ed.), *Fictionalism in metaphysics* (pp. 88–115). New York: Oxford University Press.

Yablo, S. (2010). *Things: Papers on objects, events, and properties*. Oxford/New York: Oxford University Press.

Yablo, S. (2012). Explanation, extrapolation, and existence. *Mind, 121*(484), 1007–1029.

Ylikoski and Kuorikoski. (2010). Dissecting explanatory power. *Philosophical Studies, 148*(2), 201–219.

Chapter 4
Typed Object Theory

Edward N. Zalta

Abstract The main features of typed object theory are: (1) it is formalized in a relational type theory in which relations are primitive and are not reconstructed as functions, (2) it contains a formal theory of relations, with identity conditions that allow for the hyperintensionality of relations, (3) the domain of each type t divides up into ordinary entities of type t and abstract entities of type t, and (4) the denotation and the sense of a natural language term of type t are assigned objects of the very same type, though the sense of a term of type t is always an abstract object of that type.
We show how typed object theory compares with other intensional type theories. For example, typed object theory doesn't require a primitive type for truth-values, doesn't require a primitive type for possible worlds, doesn't collapse the types for individuals and propositions (as suggested by Partee, Liefke, and Liefke and Werning); and uses only one set of types for both its syntax and semantics (unlike Williamson's logic). Finally, we analyze a number of natural language contexts without requiring the technique of 'type-raising'.

Keywords Relational type theory · Object theory · Intensionality · Hyperintensionality · Natural language semantics · Type-raising

The theory of abstract objects, hereafter 'object theory', is a system that axiomatizes both ordinary and abstract individuals, on the one hand, and properties, relations, and propositions, on the other. It has been applied in a variety of ways, for example:

I'd like to thank Paul Oppenheimer for carefully reading a draft of this paper and for his valuable suggestions for improvement.

E. N. Zalta (✉)
Center for the Study of Language and Information, Stanford University, Stanford, CA, USA
e-mail: zalta@stanford.edu

J. L. Falguera, C. Martínez-Vidal (eds.), *Abstract Objects*, Synthese Library 422,
https://doi.org/10.1007/978-3-030-38242-1_4

- in the analysis of mathematical objects and mathematical relations,[1]
- in the analysis of (a) the failures of substitution in intensional contexts, and (b) the denotations of names of fictions,[2] and
- in the analysis of possible worlds and situations, Plato's Forms, Leibnizian concepts, Frege numbers, truth-values and the logical conception of sets, and impossible worlds.[3]

Many of the above applications can be framed within second-order, quantified modal object theory. However, others require *typed object theory*, in which object theory is formulated within a background of relational type theory. Typed object theory has been developed in a number of previous works and I henceforth assume some basic familiarity with it.[4] Its most important features are: (1) there are two atomic forms of predication that may be asserted of relations and objects with the correct types, *exemplification* formulas of the form $F^n x_1 \dots x_n$ and *encoding* formulas of the form xF^1, (2) the domains of higher-order types are populated with *primitive* relations, which are axiomatized by comprehension and identity principles, (3) every type is partitioned into ordinary and abstract objects of that type, and (4) the abstract objects of each type are axiomatized by a comprehension principle which asserts, for any condition φ on properties of objects of a given type, that there is an abstract object of that type which encodes all and only the properties that satisfy φ.[5] I shall further sketch and document these and other features of typed object theory as the occasion arises below.

In this paper, typed object theory will be examined within the context of the history of *relational* type theory, including such works as Russell (1908), Carnap (1929), Orey (1959), Schütte (1960), Church (1974), and Gallin (1975). Along the way, we shall have cause to briefly compare relational type theory with functional type theory (Church 1940, Montague 1973, and others). As part of the discussion, I

---

[1]See Zalta (1983) (VI), Zalta (2000a, 2006a); Linsky and Zalta (1995), and Nodelman and Zalta (2014).

[2]See Zalta (1983) (VI), Zalta (1988a) (9–12), and Zalta (2000b).

[3]See Zalta (1993, 1997, 1999, 2000c); Pelletier and Zalta (2000); and Anderson and Zalta (2004).

[4]See Zalta (1982, 1983) (V, VI), Zalta (1988a) (9–12, Appendix), Zalta (2000a) (§3); Linsky and Zalta (1995); and Nodelman and Zalta (2014) (47–49).

[5]The key idea underlying object theory is that abstract objects *encode* rather than exemplify the properties by which we conceive of them and that such objects may encode *only* those properties and no others. As such abstract objects may be incomplete with respect to the properties they encode: there are properties F and abstract objects x such that neither xF nor $x\overline{F}$ (where $\overline{F}$ is the negation of F). But abstract objects are complete with respect to the properties they exemplify: for every property F, every abstract object x (and indeed every object whatsoever) is such either Fx or $\overline{F}x$.

The notion of encoding didn't appear in object theory *ex nihilo*. Pelletier and Zalta (2000) show how Meinwald (1992) traces a similar idea in Plato's distinction between two kinds of predication; Anderson and Zalta (2004) show how Boolos (1987) (3) finds the idea in Frege's 'two instantiation relations'; and Zalta (2006a) shows how Kripke discusses 'a confusing double usage of predication' in his Kripke (1973) [2013] lectures. The idea also appears in other works, as documented in some of the publications on object theory cited thus far.

examine some recent work in light of typed object theory. In particular, I examine Muskens (2007), Liefke (2014) and Liefke and Werning (2018), and Williamson (2013). In each case, I try to show that conclusions drawn in those works are not inevitable if one considers how the target data may be analyzed in typed object theory. I hope to show how typed object theory provides a (possibly more) natural understanding of the phenomena that these other type theories were designed to explain, and I attempt to undermine conclusions that might call into question the way typed object theory is formulated and applied.

4.1 Relational vs. Functional Type Theory

I begin with a question raised by Partee (2009 [2007], 37):

> If I am asked why we take e and t as the two basic semantic types, I am ready to acknowledge that it is in part because of tradition, and in part because doing so has worked well. … In a certain sense Montague had a third basic type, the type of possible worlds; in Gallin's Ty2 (Gallin 1975) this is explicit. But that is not essential, since on some alternatives the basic type t is taken to be the type of propositions, inherently intensional.

I'd like to put forward another answer to Partee's opening question. It starts by pointing out that e and t are the two basic semantic types only if we start with a *functional type theory* (FTT) of the kind developed by Church and Montague. By contrast, *relational type theory* (RTT) uses a single semantic type. RTT starts with a single type i, and then where $\sigma_1, \dots, \sigma_n$ are any types, $\langle \sigma_1, \dots, \sigma_n \rangle$ $(n \geq 0)$ is the derived type for relations among objects having types $\sigma_1, \dots, \sigma_n$. When $n = 0$, the type $\langle\ \rangle$ is the type for propositions. So, Partee's question is posed within a background of FTT instead of RTT.

Even starting with FTT instead of RTT, I think Partee's question has an alternative answer, namely, that it starts with the two types e and t because FTT *requires* entities (individuals) and truth values for the analysis of predication in terms of function application. This requirement traces back, I think, to the fact that Frege assumed two primitive (mutually exclusive) domains (*functions* and *objects*) and basic formulas of the form $f(x) = y$, which combine the primitive notions of *function application* and *identity*. Frege's two primitive domains classified entities into two types and in order to analyze predication, which he understood as object x *falls under* concept F, Frege had to introduce two distinguished objects, the truth values $\mathbf{T}$ and $\mathbf{F}$. Thus, a concept F could be analyzed as a function whose values are always one of the two truth values and so x *falls under* F (i.e., predication), in effect, becomes analyzed as: $F(x) = \mathbf{T}$.

Church (1940) generalized Frege's logic to allow for functions of higher type, thereby developing the first FTT. Church used α, β, and γ as variables ranging over types, and in 1940 (56), we find:

- ι and o are type symbols.
- if α and β are type symbols, then $(\alpha\beta)$ is a type symbol.

Here ι is the type for individuals, o is the type of propositions, and $(\alpha\beta)$ is the type of functions with arguments of type β and values of type α. So, for example, a property becomes a function with type $(o\iota)$, i.e., a function from individuals to truth values, and a 2-place relation between individuals has type $((o\iota)\iota)$, i.e., a function from individuals to properties. Note that Church couldn't have rested with just a single primitive type ι for individuals and the single derived type $(\alpha\beta)$. That would only yield higher-order *mappings* and not anything that would permit him to adopt Frege's analysis of predication in terms of functional application. So he added the primitive type o as the type for a truth bearer (e.g., truth values or propositions). Thus, for Church, simple predications and more complex assertions are expressions of type o, i.e., terms that denote truth values. To assert that individual x exemplifies the property F in Church's system, F must be a variable of type $(o\iota)$ and x a variable of type ι, and the expression (Fx), which represents predication, is an expression of type o (1940, 57).

Montague (1973) extended Church's typing scheme to include what appears to be a third primitive type, namely s, for possible worlds. In 1973 (256), we find Montague defining the set of *types* as the smallest set Y such that:

- $e, t \in Y$
- whenever $a, b \in Y$, $(a, b) \in Y$
- whenever $a \in Y$, $(s, a) \in Y$.

So, for Montague, (e, t) is the type for functions from individuals to truth values, i.e., the characteristic function for a set of individuals; $(e, (e, t))$ is the type for extensional relations between individuals, i.e., the characeristic function for a set of ordered pairs; $(s, (e, (e, t)))$ is the type of functions from possible world to extensional relations between individuals, i.e., intensional relations; and propositions have type (s, t), i.e., functions from possible worlds to truth values. Gallin (1975), Cresswell (1975, 1985), Thomason (1980), Fox and Lappin (2001), and Pollard (2005, 2008) all employ variations of this scheme.[6]

[6]In Gallin (1975, 58), system Ty_2 (Two-Sorted Type Theory), we find:

e, t, s $\in \mathrm{T}_2$
$\alpha, \beta \in \mathrm{T}_2$ imply $(\alpha, \beta) \in \mathrm{T}_2$.

Cf. Cresswell (1975), where there is a categorial language and semantics based on functions. In Thomason (1980) (48–9), we find:

Basic types: e, t, and p, where p is for propositions
If σ and τ are any types, then $\langle\sigma, \tau\rangle$ is a type.

Thomasons says we may "think of $\langle\sigma, \tau\rangle$ as the type of functions from the domain of type σ to that of type τ" (49). Cresswell (1985, 69) explicitly uses $\mathrm{D}_{(\tau/\sigma_1,\ldots,\sigma_n)}$ "to indicate a class of n-place functions whose domains are taken from $\mathrm{D}_{\sigma_1}, \ldots, \mathrm{D}_{\sigma_n}$, respectively, and whose range is in D_τ". Fox and Lappin (2001) (176–7) use:

Basic Types: e for individuals and Π for propositions
Exponential Types: If A, B are types, then A^B is a type.

Semanticists like Montague naturally started with the best mathematical framework they could find, and this turned out to be a version of Church's FTT. Two basic types are needed in FTT to preserve Frege's analysis of predication in terms of function application. Without the type for truth values or propositions, one can't represent the bearers of *truth*.

4.1.1 Why Relations in RTT Were Interpreted as Functions

By contrast, the bearers of truth come *for free* as 0-place relations when you formulate relational type theory (RTT). This lends it a natural elegance. RTT follows Russell's somewhat different understanding of logic, in which relations are more basic than functions. Functions become defined for Russell as 2-place relations R such that $\forall x \forall y \forall z(Rxy \& Rxz \rightarrow y = z)$. RTT doesn't need a separate primitive type for truth values since *propositions* can be analyzed as 0-place relations. The reason semanticists haven't used RTT no doubt stems from Quine's concern about the identity of intensional entities like properties and relations. The theory of relations hasn't seemed as mathematically precise as the theory of functions, since the identity conditions for relations (conceived intensionally) aren't as clearcut as those of sets or the functions definable in terms of sets. But, as we'll see, this worry doesn't apply to the version of RTT defended in this paper.

As we've noted, RTT begins with a single primitive type for individuals. This goes back to Russell (1908) (237), who took individuals, relations, and predication as basic. Though Russell didn't introduce explicit notation for types, he clearly thought of *individuals* as forming the lowest type.[7] Carnap (1929) was the first to introduce notation for types, but his notation isn't one we currently use.[8] Rather, it seems that Orey (1959) developed the now standard definition of relational types, though he didn't regard the entities in the domains of relational types as primitive relations. He defined (1959, 73) a set of *type symbols* to be the smallest set T such that:

- $\iota \in$ T, and
- if $\tau_1, \ldots, \tau_n \in$ T, then $(\tau_1 \ldots \tau_n) \in$ T $\qquad (n \geq 1)$

Here, the type A^B is a functional type. Finally, in Pollard' (2008, 273) hyperintensional type theory, there are 3 basic types (Ent, Ind, Prop) and then a variety of functional and other complex types.

[7] I shall not be discussing ramified type theory in what follows. Our attention shall be restricted entirely to the simple theory of (relational) types. See Anderson (1989) for a discussion of a ramified type theory and intensional logic.

[8] Carnap (1929, 30–32) used type `t0` for individuals, `t1` for classes of individuals, `t2` for classes of classes of individuals, and so on. He used type `t`(00) for relations among individuals, etc. There was no empty type for propositions, however type `t1` could be rewritten as `t`(0).

Note here that Orey starts with a single primitive type ι for individuals, and has a single derived type $(\tau_1 \ldots \tau_n)$, which is the type for relations that relate arguments of type $\tau_1, \ldots, \tau_n$, respectively.

However, Orey interprets expressions of type $(\tau_1 \ldots \tau_n)$ as denoting sets of n-tuples. He uses $G(\tau)$ to denote the domain of type τ, takes $G(\iota)$ to be a nonempty set, and then takes $G(\tau_1 \ldots \tau_n)$ to be some "nonvoid subset of the set of all subsets of the cartesian product of $G(\tau_1) \times \ldots \times G(\tau_n)$" (1959, 73). So, semantically, relations are interpreted as sets of n-tuples.[9]

Church contributed to RTT as well as originating FTT. He extended Orey's notation for relational types by allowing (1974, 25) the empty relational type ( ). He starts with a single primitive type i for individuals and says that where $\beta_1, \beta_2, \ldots, \beta_m$ are any types, then $(\beta_1, \beta_2, \ldots \beta_m)$ is a type. He allows for $m = 0$, and so ( ) is the type for propositions. Church agrees that one should, in some sense, take relations seriously. He says (1974, 22):

> Russell's logic must be understood intensionally if some of its significant features are to be preserved. This means in the first place that the values of the propositional-functional (for short "functional") variables are understood to be properties, in the case of singulary functional variables, or binary relations in intension in the case of binary functional variables, ...To go with this the values of Russell's propositional variables must also be taken as intensional,[...] that is, as propositions in the abstract sense rather than either sentences on the one hand or truth-values on the other.

Nevertheless, Church interprets RTT in terms of functions. Church (1974, 26) starts with two primitive domains of propositions, namely $\mathfrak{T}$ (the domain of true propositions) and $\mathfrak{F}$ (the domain of false propositions), and analyzes relations as m-ary *functions* that take values in $\mathfrak{T} \cup \mathfrak{F}$. He interprets the domain for expressions of type $(\beta_1, \beta_2, \ldots \beta_m)$ as consisting of m-ary *functions* whose m arguments have type $\beta_1, \beta_2, \ldots, \beta_m$ and whose values are (in the domain of) propositions. So, semantically, relations are understood as functions.

Gallin (1975) preserves this understanding of relations, but extends it to a language with a modal operator. Gallin first defines a set of *types* as follows (1975, 68):

> Let $\mathtt{e}$ be any symbol which is not a finite sequence. The set $\mathtt{P}$ of *predicate types* is the smallest set such that:
>
> (i) $\mathtt{e} \in \mathtt{P}$, and
> (ii) $\sigma_0, \sigma_1, \ldots, \sigma_{n-1} \in \mathtt{P}$ imply $(\sigma_0, \sigma_1, \ldots, \sigma_{n-1}) \in \mathtt{P}$.

[9] Schütte (1960, 306) uses types 0 (for individuals) and 1 (for truth values) and then introduces relational types. But he analyzes predication *syntactically* as set membership. He sets up the language (1960, 307) so that in item (1.3.3) we find:

> If $e_1, \ldots, e_n$ are expressions of types $\tau_1, \ldots, \tau_n$ and e is an expression of type $(\tau_1, \ldots, \tau_n)$, then $(e_1, \ldots, e_n \epsilon e)$ is an expression of type 1

So it is clear why Schütte interprets relational terms (predicates) and λ-expressions as denoting sets of n-tuples.

> ... Objects of type e will be individuals, and objects of type $(\sigma_0, \sigma_1, \ldots, \sigma_{n-1})$ will be relations of n arguments, of which the first is an object of type σ_0, the second an object of type σ_1, etc.

He then defines a modal language $\mathtt{ML_P}$ and reinterprets the objects of type $(\sigma_0, \sigma_1, \ldots, \sigma_{n-1})$. He says (1975, 71):

> In $\mathtt{ML_P}$, objects of type $(\sigma_0, \sigma_1, \ldots, \sigma_{n-1})$ will be predicates (relations-in-intension) of n arguments, of which the first is an object of type σ_0, the second an object of type σ_1, etc.

Gallin then defines the semantic interpretation of a predicate of type $(\sigma_0, \sigma_1, \ldots, \sigma_{n-1})$ as a function from an index (i.e., a possible world) to a set of n-tuples drawn from the appropriate types (1975, 72–3). Gallin also allowed for Henkin-style *general* models of $\mathtt{ML_P}$, and in those models, the domain of each type is some *subset* of the domain used for standard models. In other words, in a general model, the domain of type $(\sigma_0, \sigma_1, \ldots, \sigma_{n-1})$ is some *subset* of the set of all those n-tuples $\langle a_0, \ldots, a_{n-1}\rangle$ such that a_0 is an entity of type σ_0, o_1 is an entity of type $\sigma_1, \ldots$, and a_{n-1} is an entity of type σ_{n-1}. So no matter whether you consider Gallin's standard models or general models, relational predicates are still interpreted as entities that obey the principle of extensionality: they are identical when they map the same arguments (indices) to the same values (sets of n-tuples). Thus, relations are again understood semantically in terms of functions.

It is worth mentioning, at this point in the exposition, that the *intensions* discussed by Carnap, Montague, and Gallin constitute a theoretical model of the *relations-in-intensions* discussed by Russell, Church, and others. Russell and Church do not have a theory of relations-in-intensions on which they become identical when necessarily equivalent. But the intensions of Carnap, Montague, and Gallin exhibit this feature. Montague and Gallin explicitly represent intensions as functions from possible worlds to extensional entities such as truth-values, sets of individuals, or sets of n-tuples. But the primitive relations that populate the domains of relational type theory, as developed in typed object theory, are more fine-grained than intensions as conceived by Carnap, Montague, and Gallin. Such relations may be distinct even if necessarily equivalent and they may offer a better understanding of the notion of *relation-in-intension* as understood by Russell and Church.

In the remainder of this paper, we shall follow Gallin in using $\sigma_1, \sigma_2, \ldots$ as variables ranging over types. The foregoing brief history shows that the main works on RTT took relations seriously only in the syntax; semantically, they either preserved Frege's method of reducing relations to functions or took relations to be sets of n-tuples.[10] The RTT developed in Zalta (1982, 1983, 1988a,b) stands

[10] van Benthem and Doets (1983), and Anderson (1989), followed Orey in this regard. van Benthem and Doets define (1983, 269):

$$D_{(\tau_1,\ldots,\tau_n)}(A) = \mathcal{P}(D_{\tau_1}(A) \times \cdots \times D_{\tau_n}(A))$$

i.e., they define the domain of relations among entities with types $\tau_1, \ldots, \tau_n$ to be the set containing all the sets of n-tuples with elements drawn, respectively, from the domains of type $\tau_1, \ldots \tau_n$. Muskens also interprets relational expressions as sets of n-tuples rather than as functions. See in contrast: in these works, the members of the domain for each derived type are *primitive relations*. Though the notation for types used in 1982, 1983, and 1988b was not as elegant as the notation used in 1988a, nevertheless, in all four works, no attempt was made to reduce relations to functions or sets in the semantics.[11] Using the more elegant formulation of Zalta (1988a) (231), a *type* was defined as follows:

- 'i' is a *type*
- Whenever $\sigma_1, \ldots, \sigma_n$ are any *types*, $\ulcorner\langle\sigma_1, \ldots, \sigma_n\rangle\urcorner$ is a *type* $\quad (n \geq 0)$

In the usual way, i is the type for *individuals* and $\langle\sigma_1, \ldots, \sigma_n\rangle$ is the type for *n-place relations*, the arguments of which have types $\sigma_1, \ldots, \sigma_n$, respectively. When $n = 0$, $\langle\,\rangle$ is the type for *propositions*. Thus, if we use the expression F as a typical term of type $\langle\sigma_1, \ldots, \sigma_n\rangle$ and use the expressions $x_1, \ldots, x_n$ as typical terms with types $\sigma_1, \ldots, \sigma_n$, respectively the language of typed object theory includes (atomic) *exemplification* formulas of the form $Fx_1 \ldots x_n$. And where F has type $\langle\sigma\rangle$ and x has type σ, the language includes (atomic) *encoding* formulas of the form xF.

In all the works on typed object theory, however, the semantics included, for each type σ, a non-empty domain $\mathcal{D}_\sigma$ consisting of primitive entities, as well as a separate non-empty domain $\mathcal{W}$ of possible worlds. Relations in the domains of type $\mathcal{D}_{\langle\sigma_1,\ldots,\sigma_n\rangle}$ were then clearly distinguished from set-theoretic representations thereof by introducing a world-relative *exemplification extension* function that essentially maps each relation to a Montagovian intension. Specifically, this function mapped each pair consisting of an n-place primitive relation $\boldsymbol{r}$ in $\mathcal{D}_{\langle\sigma_1,\ldots,\sigma_n\rangle}$ and a possible world $\boldsymbol{w}$ in $\mathcal{W}$ to a set of n-tuples drawn from the power set of $\mathcal{D}_1 \times \ldots \times \mathcal{D}_n$.[12] So the exemplification extension function helps one give a semantic representation of the world-relative truth conditions of atomic exemplification formulas: where F has type $\langle\sigma_1, \ldots, \sigma_n\rangle$ and $x_1, \ldots, x_n$ have types $\sigma_1, \ldots, \sigma_n$, respectively, the formula '$Fx_1 \ldots x_n$' is true at a world w just in case the n-tuple of objects denoted

---

his definition of Orey frames (Muskens 1989, 2). But later in this paper, in his system $\mathrm{TT}^{\eta,2}$, we discover that he doesn't take *predications* as basic formulas of the language. Instead, function application is basic (1989, 12, Definition 10, ii), as it is in Muskens (1995) (p. 15, Definition 9, iv).

[11] In 1982 (298), and 1983 (109), the relational types were defined basically as follows:

- i is a *type*
- p is a *type*
- Whenever $\sigma_1, \ldots, \sigma_n$ are any types, $\ulcorner(\sigma_1, \ldots, \sigma_n)/p\urcorner$ is a type $\;(n \geq 1)$

So in these early works, $(\sigma_1, \ldots, \sigma_n)/p$ is the derived type for relations. Since the primitive type p in these works corresponds to the new derived type $\langle\,\rangle$ in 1988a, the structure of the types is the same. But whereas Zalta (1988a) clearly uses a single primitive type, Zalta (1982, 1983, 1988b) suggest that there are two primitive types (see, e.g., Zalta 1988b, 69). In these latter works, I was still under the influence of the Montague Grammar I had learned as a graduate student, and hadn't yet recognized that one could eliminate p as a second primitive type by defining type p as the empty derived type $\langle\,\rangle$.

[12] See Zalta (1982) (299); Zalta (1983) (114); and Zalta (1988a) (236). The exemplification function also maps each proposition $\boldsymbol{r}$ in $\mathcal{D}_{\langle\,\rangle}$ and a possible world $\boldsymbol{w}$ in $\mathcal{W}$ to a truth value.

by $x_1, \ldots, x_n$ is an element of the exemplification extension at $\boldsymbol{w}$ of the relation denoted by F.

Thus, the relational predicates of typed object theory *do not denote* Montagovian intensions; the semantics allows for distinct relations that have the same exemplification extension at every possible world. The semantics of typed object theory also includes an *encoding extension* function that maps each property in each domain $\mathcal{D}_{\langle\sigma\rangle}$ to a set of objects drawn from the domain $\mathcal{D}_\sigma$. The encoding extension function is used to give the truth conditions for atomic encoding formulas: where x has type σ and F has type $\langle\sigma\rangle$, then the encoding formula 'xF' is true at a world $\boldsymbol{w}$ just in case the object in $\mathcal{D}_\sigma$ denoted by x is an element of the *encoding extension* of the property in $\mathcal{D}_{\langle\sigma\rangle}$ denoted by F. Since the truth conditions of 'xF' are independent of the possible worlds, the formula $xF \rightarrow \Box xF$ will be valid.

The reason typed object theory doesn't follow the tradition of interpreting the relations in $\mathcal{D}_{\langle\sigma_1,\ldots,\sigma_n\rangle}$ as functions or sets is that it comes with a precise theory of relations. With a precise theory of relations in hand, one can argue that *no mathematical model* of relations in set theory is needed to represent them in the semantics. Indeed, one shouldn't use a mathematical model that collapses necessarily equivalent relations, given that object theory doesn't require the collapse. This stands in contrast to the versions of RTT in Orey, Church, and Gallin. Each of these yields artifactually valid statements, i.e., statements whose validity is required by the semantic representation but not required given the nature of the objects being represented. For example, either $\forall x(Fx \equiv Gx) \rightarrow F = G$ or $\Box\forall x(Fx \equiv Gx) \rightarrow F = G$ is artifactually valid in other versions of RTT. But object theory isn't committed to either claim.

Before we discuss how typed object theory approaches identity for relations, we first review how identity for relations is defined in *second-order*, modal object theory. In second-order object theory, one can state a precise comprehension (i.e., existence) schema as well as identity conditions for relations. The latter offer *extensional* conditions for the identity of relations, notwithstanding our conception of them as intensional entities. Readers familiar with the second-order version of object theory will recall that the language includes two modes of predication, *exemplification* formulas of the form $\Pi\kappa_1 \ldots \kappa_n$ (where Π is any n-place relation term and $\kappa_1, \ldots, \kappa_n$ are any individual terms), and *encoding* formulas of the form $\kappa\Pi$ (where κ is any individual term and Π any 1-place relation term, i.e., any property term). In the second-order version of object theory, properties F and G are semantically assigned not only an exemplification extension that can vary from world to world, but also an *encoding* extension. Thus, the language of second-order object theory distinguishes the following conditions on properties F and G:

(A) $\Box\forall x(Fx \equiv Gx)$
(B) $\Box\forall x(xF \equiv xG)$

We may then reject the idea that (A) provides identity conditions for F and G. But we accept (B) as providing correct identity conditions for properties even when we conceive of properties as intensional entities. That is, we define:

$$F = G \ =_{df} \Box\forall x(xF \equiv xG)$$

Moreover, given the logical principle that $xF \rightarrow \Box xF$, we may infer $\Box\forall x(xF \equiv xG)$ whenever $\forall x(xF \equiv xG)$. Hence, to prove that properties are identical in object theory, one need only prove $\forall x(xF \equiv xG)$. Thus, the foregoing definition and the logic of encoding gives us extensional identity conditions for intensional entities, since they intuitively imply that properties are identical when their encoding extensions are identical.

Moreover, in second-order object theory, identity conditions for propositions and for n-place relations ($n \geq 2$) can be defined in terms of the definition of property identity. Consider propositions first. Let x be a variable ranging over individuals, and let p and q be variables ranging over propositions. Thus, in second-order object theory, expressions such as $[\lambda x\ p]$ ('*being such that* p') , $[\lambda x\ q]$ ('*being such that* q'), etc., denote properties of individuals; the properties denoted depend on the value assigned to p, q, etc. So we may say:

> Propositions p and q are identical if and only if the properties *being such that p* and *being such that q* are identical, i.e.,
>
> $$p = q \ =_{df} \ [\lambda x\ p] = [\lambda x\ q]$$

Here, proposition identity has been reduced to property identity. Now, for n-place relations where $n \geq 2$, second-order object theory allows us to define identity as follows. Let F and G both be n-place relation variables, for some $n \geq 2$. Then, we may say:

> F and G are identical just in case every way of 'plugging' $n - 1$ individuals into F and G (plugging them in the same order into F and G) results in identical properties.

We can make this precise by using the following formal definition, which is well-formed in second-order object theory, where the arity of both F and G is some n such that $n \geq 2$:

$$\begin{aligned} F = G \ =_{df} \ & \forall x_1 \ldots \forall x_{n-1}([\lambda y\ Fyx_1 \ldots x_{n-1}] = [\lambda y\ Gyx_1 \ldots x_{n-1}]\ \& \\ & [\lambda y\ Fx_1yx_2 \ldots x_{n-1}] = [\lambda y\ Gx_1yx_2 \ldots x_{n-1}]\ \&\ \ldots\ \& \\ & [\lambda y\ Fx_1 \ldots x_{n-1}y] = [\lambda y\ Gx_1 \ldots x_{n-1}y]) \end{aligned}$$

These definitions make it clear that second-order object theory doesn't automatically collapse properties, relations, and propositions that are necessarily equivalent in the classical sense.

In particular, second-order object theory doesn't collapse properties that are necessarily equivalent in the sense of (A). There are lots of examples of properties F and G that are distinct despite being necessarily exemplified by the same objects. For example, the property *being a barber who shaves all and only those who don't shave themselves*, i.e.,

$$[\lambda x\ Bx\ \&\ \forall y(Sxy \equiv \neg Syy)]$$

is clearly distinct from the property *being a dog that is both white and not white*:

$$[\lambda x\ Dx\ \&\ Wx\ \&\neg Wx]$$

While these properties clearly satisfy (A), we may consistently assert that they are distinct properties. As we shall see below, this will imply that there are abstract objects that encode one without encoding the other. Second-order object theory, however, stipulates that properties that satisfy (B) are identical. Examples of properties that are identical include: *being a brother* and *being a male sibling*, *being a circle* and *being a closed plane figure every point of which lies equidistant from some given point*, etc.

These features of second-order, modal object theory are, in part, a consequence of the comprehension principle for abstract individuals asserted as part of that theory. This principle is formulated as follows, where $A!$ denotes the property of *being abstract*:

$$\exists x(A!x\ \&\ \forall F(xF \equiv \varphi)),\ \text{provided } x \text{ doesn't occur free in } \varphi$$

In other words, for any condition φ on properties F, there is an abstract object that encodes exactly those properties F such that φ. Thus, once we add the assertion that, for some given property pair P and Q that $P \neq Q$, it follows that there is an abstract object that encodes the one without encoding the other.[13] Moreover, if F and G are identical, there couldn't be an abstract object that encodes one without encoding the other.

4.1.2 Identity in Typed Object Theory

Typed object theory, in contrast to second-order, modal object theory, gives us more flexibility in defining identity. In the discussions of typed-object theory in Zalta (1983, 1988a), the definitions for relation identity essentially followed the foregoing discussion. As a result, the definitions in those works were more 'type-specific' in that they defined identity for type i one way and defined identity for all relational types using principles analogous to those used in the second-order case.[14] However,

[13]By comprehension, we have, using the defined notion of identity:

$$\exists x(A!x\ \&\ \forall F(xF \equiv F = P))$$

Clearly, this object encodes P without encoding Q, given $P \neq Q$.

[14]In Zalta (1983) (121, 124) and Zalta (1988a) (241–2), we first defined identity for individuals as follows, where x and y are variables of type i, F is a variable of type $\langle i \rangle$, and $O!$ and $A!$ are the predicates of *being ordinary* and *being abstract*, respectively, both of type $\langle i \rangle$:

in what follows, we adopt the definitions in Zalta (1982) (301–2) and (2000a) (228), where we find identity defined for all types using a single formula scheme. First, recall that the domain of every type σ is partitioned into *ordinary* and *abstract* objects of that type. We use the typically ambiguous predicates $O!^{\langle\sigma\rangle}$ and $A!^{\langle\sigma\rangle}$, for every type σ, to denote the properties of *being ordinary* and *being abstract*, respectively.[15] Consequently, there are both ordinary and abstract properties of individuals, ordinary and abstract relations among individuals, etc.[16]

It is axiomatic that ordinary objects of type σ only exemplify properties, whereas abstract objects of type σ both exemplify and encode properties. Moreover, the abstract objects of each type σ are governed by a typed version of the comprehension principle discussed earlier. The typed version of this principle can be stated generally as follows. Let x be a variable of any type σ, F be a variable of type $\langle\sigma\rangle$, and $A!$ be a predicate (mentioned earlier) of type $\langle\sigma\rangle$. Then we take the instances of the following to be axioms, for any type σ:

$$\exists x(A!x \;\&\; \forall F(xF \equiv \varphi)),\ \text{provided } x \text{ doesn't occur free in } \varphi$$

In other words, for any condition φ that places a condition on properties of type-σ objects, there exists an abstract object x of type σ that encodes all and only the properties F such that φ.

Now since the domain of each type is partitioned into ordinary and abstract objects of that type, we can state identity conditions for objects of any type σ as follows. Let x and y be variables of any type σ; let F be a variable of type $\langle\sigma\rangle$; and let $O!$ and $A!$ be the predicates of *being ordinary* and *being abstract*, respectively, each having type $\langle\sigma\rangle$. Then we define:

$$x = y \;=_{df}\; (O!x \;\&\; O!y \;\&\; \Box\forall F(Fx \equiv Fy)) \vee (A!x \;\&\; A!y \;\&\; \Box\forall F(xF \equiv yF))$$

Then, for any type σ, let x is a variable of type σ, and let F and G be variables of type $\langle\sigma\rangle$. Then we defined:

$$F = G \;=_{df}\; \Box\forall x(xF \equiv xG)$$

Finally, identity for propositions and n-place relational types were defined along the lines used for second-order object theory. See the references cited at the beginning of this note for further details.

[15] Actually, *being ordinary* and *being abstract* are defined technical terms in typed object theory; we used a typically-ambiguous primitive predicate, $E!^{\langle\sigma\rangle}$ (for every σ), where this predicate intuitively picks out the concrete objects of type σ. Now let x be a variable of type σ, $O!$ be the predicate for *being ordinary*, with type $\langle\sigma\rangle$, and $A!$ be the predicate for *being abstract*, also with type $\langle\sigma\rangle$. Then we may say x exemplifies *being ordinary*, written $O!x$, just in case $\Diamond E!x$, and x exemplifies *being abstract*, written $A!x$, just in case $\neg\Diamond E!x$. This holds for every type σ. Thus, the domain of every type is partitioned into the ordinary and abstract objects of that type.

[16] The standard primary and secondary qualities count as good examples of ordinary properties of individuals, and we may include empty properties, such as *being a giraffe in the Arctic Circle*, *being round and square*, as ordinary properties of individuals. But fictional properties (e.g., *being a hobbit*, *being composed of phlogiston*, etc.) and mathematical properties (e.g., *being a Peano number*, *being a ZF set*, etc.) have been analyzed as abstract properties. These latter encode just the properties of properties attributed to them in their respective story or theory. *Absolute simultaneity*, the *membership relation* of ZF, etc., are examples of abstract relations.

(C) $x = y =_{df} (O!x \;\&\; O!y \;\&\; \Box\forall F(Fx \equiv Fy)) \vee (A!x \;\&\; A!y \;\&\; \Box\forall F(xF \equiv yF))$

In other words, ordinary objects of any type σ are identical whenever they necessarily exemplify the same type-$\langle\sigma\rangle$ properties, and abstract objects of type σ are identical whenever they necessarily encode the same type-$\langle\sigma\rangle$ properties. Clearly, given the theorem $O!x \vee A!x$, we can derive $x = x$ from (C).[17] So, by taking the substitution of identicals as an axiom, typed object theory has a theory of identity in which substitution of identicals holds in any context.

To see an example of (C) in action, fix σ and let F and G be variables of type $\langle\sigma\rangle$; let $O!$ and $A!$ be the predicates *being ordinary* and *being abstract* with type $\langle\langle\sigma\rangle\rangle$; and let $\mathcal{F}$ be a variable of type $\langle\langle\sigma\rangle\rangle$. Then the following is an instance of (C) (where we've reduced the font size of 'F' and 'G' for readability):

(D) $F = G =_{df} (O!F \,\&\, O!G \,\&\, \Box\forall\mathcal{F}(\mathcal{F}F \equiv \mathcal{F}G)) \vee (A!F \,\&\, A!G \,\&\, \Box\forall\mathcal{F}(F\mathcal{F} \equiv G\mathcal{F}))$

This definition governs any properties F and G with type $\langle\sigma\rangle$, for any σ. And, clearly, it yields the theorem $F = F$.

Note that (D) still allows necessarily equivalent properties to be distinct even when necessarily equivalent, i.e., (C) doesn't imply the typed version of (A). Moreover, the definition of property identity used in second-order object theory now becomes a theorem. That is, where x is a variable of type σ, and F and G have the types assigned in the previous paragraph, we may prove:

$$F = G \equiv \Box\forall x(xF \equiv xG)$$

We leave the proof to a footnote.[18] Interestingly, the status, in typed object theory, of the second-order definitions proposition identity and n-place relation identity ($n \geq$

[17] This follows by disjunction syllogism from the theorem $O!x \vee A!x$. Suppose $O!x$. Then since it is a modal theorem that $\Box\forall F(Fx \equiv Fx)$, we have $O!x \;\&\; O!x \;\&\; \Box\forall F(Fx \equiv Fx)$. By $\vee$I, this gives us that right-side of (C). So $x = x$. On the other hand, suppose $O!x$. Then since it is a modal theorem that $\Box\forall F(xF \equiv xF)$, we have $A!x \;\&\; A!x \;\&\; \Box\forall F(xF \equiv xF)$. By $\vee$I, this gives us that right-side of (C). So $x = x$.

[18] ($\rightarrow$) This direction is trivial, by the substitution of identicals and the modal theorem $\Box\forall x(xF \equiv xF)$. ($\leftarrow$). Assume $\Box\forall x(xF \equiv xG)$. Then by the T schema, $\forall x(xF \equiv xG)$. Now where $A!$ is the predicate *being abstract* of type $\langle\sigma\rangle$ and H is a variable of the same type, then using the defined notion of identity in (D), we have the following instance for comprehension for abstract individuals:

$$\exists x(A!x \;\&\; \forall H(xH \equiv H = F))$$

Call such an individual a, so that we know both $A!a$ and $\forall H(aH \equiv H = F)$. Instantiating the latter to F and G, respectively, we have both:

(ϑ) $\quad aF \equiv F = F$
(ξ) $\quad aG \equiv G = F$

Since $F = F$, it follows from (ϑ) that aF. But, it follows from a previously established fact, namely $\forall x(xF \equiv xG)$, that $aF \equiv aG$. Hence aG. Now, for reductio, assume $F \neq G$, i.e., by symmetry of identity, $G \neq F$. Then by (ξ), $\neg aG$. Contradiction. $\bowtie$.

2) is a subject of ongoing investigation. But even though their status isn't settled, we still have a fully general theory of identity that applies to every type σ, namely (C).

With a precise theory of relations (i.e., existence and identity conditions) in hand, we have a foundational framework that allows one to represent classical predication in relational terms. We can regard 'x loves y', 'x worships y', '$x \in y$', '$x \leq y$', etc., as exemplification predications of the form Rxy, where x and y are individuals and R is a relation of type $\langle i, i\rangle$. From the point of view of typed object theory, there is nothing more fundamental than individuals and relations. We have no need of (a) the Fregean tradition of interpreting relations as functions and interpreting predication as functional application, or (b) the set-theoretic tradition of interpreting relations as sets of n-tuples and interpreting predication as set membership.[19] Such traditions *fail to capture* the essential fact about predication, namely, that in a true exemplification predication of the form Fx, the property F *characterizes* x; it doesn't merely classify x or correlate (i.e., map) x to a truth-value.

For the remainder of this paper, we therefore assume it is a mistake to regard relations as functions or sets; such an interpretation collapses necessarily equivalent relations and validates principles to which typed object theory is not committed. Instead, general Henkin models in which the domain of each type consists of primitive entities of that type gives a more accurate picture of the ontology that underlies RTT in general and typed object theory in particular.[20] Moreover, as we shall see, typed object theory has some theoretical virtues when compared to other recent intensional interpretations of RTT.

4.2 Intensional Type Theory: I

It may be of interest to see what typed object theory accomplishes when compared to the framework developed in Muskens (2007). Muskens states clearly that he is not so much interested in the question of what the intensional entities that populate the domains of RTT *are*, but rather interested in the general features of any RTT in which the relational types denote *intensions* that can be distinguished from extensional entities such as sets of n-tuples. He traces a two-stage pattern of

[19] See Bueno et al. (2014) for a discussion of how the theory of propositions and possible worlds is completely 'set free' in object theory.

[20] From the present standpoint, the semantics of the language of typed object doesn't provide any further theoretical understanding of the primitive relations that populate the domains $\mathcal{D}_\sigma$, for $\sigma \neq i$. Indeed, metaphysically, the language of set theory used in a typical model-theoretic semantics can be analyzed within typed object theory. But, of course, if we allow ourselves some set theory and urelements, we can develop a model-theoretic semantics for the language of typed object theory. See Zalta (1983, 1988a).

semantically distinguishing intensions and extensions back to Frege's distinction between sense and reference. He writes (2007, 101):

> Thus, while opinions about the nature of intensions radically diverge, all proposals follow a simple two-stage pattern. The aim of this paper is not to add one more theory of intension to the proposals that have already been made, but is an investigation of their common underlying logic. The idea will be that the two-stage set-up is essentially all that is needed to obtain intensionality. For the purposes of logic it suffices to consider intensions as abstract objects; the question what intensions are, while philosophically important, can be abstracted from.

His general system for studying RTT under intensional interpretations, which he calls ITL ('Intensional Type Logic'), has some very interesting properties.

One basic difference between ITL and typed object theory concerns the language. ITL doesn't have a primitive form of *predication*, whereas typed object theory has two. Instead, ITL has primitive function application of the form (AB); where A is a relational term of type $\langle \sigma_1, \ldots, \sigma_n \rangle$ and B is a term of type σ_1, (AB) is a term of type $\langle \sigma_2, \ldots, \sigma_n \rangle$. So, in effect, ITL doesn't treat statements of the form x loves y, $x \in y$, $x \leq y$, etc., as instances of the *primitive* form of predication Rxy. Instead, the semantics shows that relations are treated functionally. Muskens explains the semantic clause that assigns a value to terms of the form (AB) as follows[21]:

> To better understand the motivation behind the second and third clauses of this definition, it may help to consider that any $n + 1$ place relation R can be thought of as a unary function F such that $F(d) = \{ \langle \vec{d} \rangle \mid \langle d, \vec{d} \rangle \in R\}$.

So it seems clear that Muskens, like Orey, Church, and Gallin, is not taking relations in RTT as primitive entities.

Putting this aside, the general logical framework Muskens develops for RTT has a number of virtues. As he explains, the framework allows us to distinguish formulas that are 'co-entailing' (2007, 113), allows us to represent the sense/reference distinction (2007, 114), and allows for a construction of possible worlds (2007, 115).

However, I think that typed object theory offers somewhat more general analyses of the same phenomena precisely because it provides a theory of the intensions that populate the domains of relational types. First, as noted earlier, typed object theory distinguishes properties and relations from their Montagovian intensions. Let us return to the examples of the two properties that are distinct but necessarily equivalent: $[\lambda x\, Bx \,\&\, \forall y(Sxy \equiv \neg Syy)]$ ('being a barber who shaves all and only those who don't shave themselves') is a property of type $\langle i \rangle$, as is $[\lambda x\, Dx \,\&\, Wx \,\&\neg Wx]$

---

[21]The semantic clause in question is (2007, 104):

$$V(a, AB) = \{ \langle \vec{d} \rangle \mid \langle I(a, B), \vec{d} \rangle \in V(a, A) \}$$

Since $V(a, X)$ is generally defined to be the extension of the intension of X, I would gloss the above as follows: the extension of the intension of (AB) is the set of $n - 1$ tuples obtained by removing the first member of each n-tuple in the extension of the intension of A whose first member is the intension of B.

('being a dog that is white and non-white'). In these expressions, 'B', 'D', and 'W' are of type $\langle i \rangle$, and 'S' is of type $\langle i, i \rangle$. Now let:

- s ('Sally') denote an individual of type i,
- j ('John') denote an individual of type i,
- $Bel(\ ,\ )$ ('believes') denote a relation of type $\langle i, \langle\ \rangle \rangle$,
- $[\lambda\ \varphi]$ ('that φ') denote a proposition of type $\langle\ \rangle$, provided φ has no encoding subformulas

Then, in typed object theory, the claim:

> Sally believes that John is a barber who shaves all and only those who don't shave themselves, i.e.,
>
> $$Bel(s, [\lambda\ [\lambda x\ Bx\ \&\ \forall y(Sxy \equiv \neg Syy)]j])$$

doesn't imply:

> Sally believes that John is a dog that is both white and not white, i.e.,
>
> $$Bel(s, [\lambda\ [\lambda x\ Dx\ \&\ Wx\ \&\neg Wx]j])$$

Typed object theory has had this feature from its inception, but with principles that articulate existence and identity conditions for the intensions that play a role in the above representations, namely, relations Bel and S, properties B, D, and W, and propositions $[\lambda\ [\lambda x\ Bx\ \&\ \forall y(Sxy \equiv \neg Syy)]j]$ and $[\lambda\ [\lambda x\ Dx\ \&\ Wx\ \&\neg Wx]j]$.

Moreover, to avoid the problems of hyperintensionality, which requires us to explain why the fact that John believes that Cicero is a Roman doesn't imply that John believes that Tully is a Roman, Muskens has to interpret proper names as (higher-order) properties of properties, and justify this by appeal to a principle of the 'primacy of properties' (2007, 114). By contrast, typed object theory treats 'Cicero' and 'Tully' as names that denote individuals. It then uses a completely general analysis of Fregean senses to explain the problems of hyperintensionality. We sketch this analysis briefly.

In typed object theory, the sense of a natural language expression with type σ isn't an entity of higher type. Rather the sense is of the very same type. The sense of an expression of type σ is an abstract object of type σ, i.e., one that encodes properties with type $\langle \sigma \rangle$. By encoding properties of σ-type objects, the sense can *represent* an object that exemplifies the properties in question, though object theory doesn't require that sense determines reference! Indeed, object theory allows the sense of an expression to vary from person to person and that for many expressions, the sense of that expression for a person can encode properties that the object denoted by the expression fails to exemplify. But, in what follows, we suppress this feature of the theory.

Consider type i expressions 'Samuel Clemens' ('c') and 'Mark Twain' ('t'), which are learned in different contexts. The denotation (extension) and the senses (intensions) of 'c' and 't' are of type i:

- 'c' and 't' denote the same ordinary individual
- The sense of 'c' and the sense of 't' are distinct abstract individuals.

In typed object theory, we may represent the sense of 'c' and 't' as '$\underline{c}$' and '$\underline{t}$', respectively.[22] We then have a way to model Frege's solution to the problem of cognitive significance of identity statements for proper names. 'Cicero is Cicero' ($c = c$) is knowable a priori, whereas 'Cicero is Tully' ('$c = t$') is true but not knowable a priori: 'Cicero' and 'Tully' are expressions that have the same denotation (namely, Cicero) and different senses (namely $\underline{c}$ and $\underline{t}$).

Moreover, this analysis generalizes to all higher types. The sense of an expression of type $\langle \sigma_1, \ldots, \sigma_n \rangle$ is an abstract object of that very type. There is no type-raising. Consider, for example, the type $\langle i \rangle$ expressions 'woodchuck' ('W') and 'groundhog' ('G'), which are learned in different contexts. The denotation (extension) and the senses (intensions) of 'W' and 'G' are of type $\langle i \rangle$:

- 'W' and 'G' denote the same property of individuals
- The sense of 'W' and the sense of 'G' are distinct abstract properties – they encode different properties of properties.

We may represent the sense of 'W' and 'G' in typed object theory as '$\underline{W}$' and '$\underline{G}$', respectively, again suppressing possible indices to persons, times or contexts. This provides a Fregean solution to the problem of the cognitive significance of identities: whereas 'being a woodchuck is identical to being a woodchuck' ('$W = W$') is knowable a priori, 'being a woodchuck is identical to being a groundhog' ('$W = G$') is not; the expressions 'being a woodchuck' and 'being a groundhog' have the same denotation but different senses.

Thus, we may explain hyperintensionality both at the level of individuals and at every higher type. At the level of individuals, we represent belief reports in terms of an ambiguity: the expressions 'Cicero' and 'Tully' contribute their denotations on the *de re* readings, but contribute their senses on the *de dicto* readings:

- John believes that Cicero is a Roman.

  (1) $B(j, [\lambda\ Rc])$ (*de re*)
  (2) $B(j, [\lambda\ R\underline{c}])$ (*de dicto*)

- John doesn't believe that Tully is a Roman.

  (3) $\neg B(j, [\lambda\ Rt])$ (*de re*)
  (4) $\neg B(j, [\lambda\ R\underline{t}])$ (*de dicto*)

- Cicero is Tully.

  (5) $c = t$

[22] These can be indexed to persons and times or contexts, but as noted previously, we'll omit this relativization. The important point is that, as abstract objects, $\underline{c}$ and $\underline{t}$ can encode properties that Cicero exemplifies. (Or not, if one really wants to have a better understanding of how language works, as opposed to simply following Frege in discussing an ideal language.)

We explain the hyperintensionality by the fact that the *de dicto* readings (2) and (4) are consistent, even given (5).

The very same explanation can be given for hyperintensionality of the woodchuck/groundhog case. The expressions ‘woodchuck’ and ‘groundhog’ contribute their denotations on the *de re* readings, but contribute their senses on the *de dicto* readings:

- John believes that Woody is a woodchuck.

  (6) $B(j, [\lambda\, Ww)$ (*de re*)
  (7) $B(j, [\lambda\, \underline{W}w])$ (*de dicto*)

- John doesn’t believe that Woody is a groundhog.

  (8) $\neg B(j, [\lambda\, Gw])$ (*de re*)
  (9) $\neg B(j, [\lambda\, \underline{G}w])$ (*de dicto*)

- Being a woodchuck just is being a groundhog.

  (10) $W = G$

Again, the *de dicto* readings (7) and (9) are consistent, even given (10). Note also that object theory even offers the reading $B(j, [\lambda\, \underline{W}\, \underline{w}])$, in which both the sense of the individual term and the sense of the predicate are combined in the proposition that is the object of belief. The consequences of these readings were developed in other works on object theory; see Zalta (1988a) (166–172); and Zalta (2001) (337–341).

One final point of comparison with (Muskens 2007) is in order. Muskens suggests that possible worlds can be constructed in ITL as properties of propositions.[23] Using the ITL variables w of type $\langle\langle\rangle\rangle$ (i.e., the type for properties of propositions) to range over possible worlds, he extends the system to include (2007, 115):

- a new primitive predicate Ω (‘is a world’) with type $\langle\langle\langle\,\rangle\rangle\rangle$,
- axioms that assert (1) if w is a world, then the false proposition ($\bot$) is not true at w, and (2) if w is a world, then if a conditional $A \rightarrow B$ is true at w, then for any objects, if it is true at w that A characterizes those objects, then it is true at w that B does too,
- a new primitive constant w_0 to designate the *actual world*, and
- axioms that govern w_0, which stipulate that w_0 is a possible world and that all and only true propositions are true at w_0

[23] Some of the following observations, suitably adjusted, apply to the reconstruction of possible worlds in the FTTs articulated in Fox and Lappin (2001) (184–7), and Pollard (2005) (41–3), Pollard (2008) (276–7). These authors assume a domain of primitive propositions structured as a pre-Boolean algebra or prelattice and then define possible worlds as ultrafilters (maximal prime filters) on this domain. This is clearly a *model* of possible worlds and *truth at a world*, not a theory of these notions. The propositions in a set do not characterize the set in any way. By contrast, possible worlds that *encode* propositional properties are characterized by these properties, since encoding is a mode of predication.

These basic features come with the nice feature that the notion, proposition p *is true at* world w, is just defined as wp (i.e., the result of applying w to p).

By constrast, object theory doesn't need a new primitive predicate for possible worlds. Possible worlds are defined as *situations*, which are in turn defined as abstract individuals that encode propositions by encoding only propositional properties (Zalta 1983 (IV); Zalta 1993; and Menzel and Zalta 2014). A *possible world* is defined as any abstract individuals that might be such that it encodes all and only true propositions. Moreover, *truth at a world* is defined in terms of encoding: p is true at w (written $w \models p$) if and only if w encodes *being such that* p, i.e., if and only if $w[\lambda y\ p]$. An actual world is then defined to be a possible world w such that $\forall p((w \models p) \equiv p)$. I won't rehearse these definitions in detail here, but merely assert that the *axioms* Muskens asserts to construct possible worlds are *theorems* of object theory. It is provable in object theory that: (1) no contradiction is true at any world, (2) that if $w \models (p \rightarrow q)$ and $w \models p$, then $w \models q$, (3) every world is maximal (i.e., for any w and any proposition p, either $(w \models p) \vee (w \models \neg p)$, and (4) there is a unique actual world (see, e.g., Zalta 1993).

Moreover, object theory is developed in a modal setting. So, its theory of worlds also yields the following claims as theorems:

$$\forall p(\Box p \equiv \forall w(w \models p))$$

$$\forall p(\Diamond p \equiv \exists w(w \models p))$$

Thus, the object-theoretic analysis of worlds implies the fundamental facts about possible worlds as theorems: a proposition is necessarily true if and only if it is true in all possible worlds, and proposition is possibly true if and only if it is true in some possible world. These principles draw a deep connection between our pre-theoretical understanding of necessity and possibility and our theoretical understanding of possible worlds. With such principles as theorems, all we have to do to prove the existence of non-actual possible worlds is to assert, for some proposition p, that $\neg p\ \&\ \Diamond p$, for it then follows that there exists a possible world distinct from the actual world where p is true. It is not clear whether this connection between our pre-theoretic understanding of modality and our theoretical understanding of possible worlds is preserved by the analysis of the modal operators we find in Muskens (2007) (116).[24]

[24] I say this because Muskens has to stipulate the axioms he labels W3 and W4, which assert that when *truth at a world* and *being a world* hold of the appropriate objects, they hold by necessity. By constrast, object theory yields these claims as theorems: one can prove in object theory that $(w \models p) \rightarrow \Box(w \models p)$ and that $\mathit{PossibleWorld}(w) \rightarrow \Box\mathit{PossibleWorld}(w)$. These facts hold because both the notions of *truth at a world* and *possible world* are defined in terms of encoding formulas, which are governed by the axiom $xF \rightarrow \Box xF$. Also, it looks like the analysis Muskens offers (2007, 116) has to build the fundamental principles connecting modality and *truth at* a world into the R (accessibility) relation, so that principles like the ones being discussed in the text end

I conclude this section with two further observations. The first is that object theory doesn't need any special new axioms to develop the theory of *impossible* worlds: an impossible world i is a situation that is *maximal* and such that it is not possible that every proposition true in i is true, i.e., $\neg\Diamond\forall p((i \models p) \rightarrow p)$. So, in the special cases of hyperintensionality where impossible worlds are needed (e.g., for counterfactuals with impossible antecedents), the theory provides the background theoretical entities needed for the analysis to proceed.

Second, there may be an issue with Muskens' reconstruction of possible worlds. If worlds are, as he says, properties of propositions, and properties are intensional entities, then he may have *too many* possible worlds. This is clearest in the case of the actual world w_0. If w_0 is a property of propositions, then consider any property of propositions that is distinct from w_0 but necessarily equivalent to it. Then we would have *two* distinct actual worlds. In other words, his definitions and axioms don't guarantee that there exists a unique actual world. By contrast, in object theory, it is provable that there is a unique actual world (i.e., there is a unique abstract object that is a possible world and is actual, namely, the abstract object that encodes all and only the properties of the form $[\lambda y\ p]$, where p is a true proposition. The problem of *too many worlds* affects other well-known attempts to define possible worlds as fine-grained intensional entities such as *states of affairs* (see Zalta 1988a, 72–74, for further discussion).

4.3 Nominalized Propositions

Recently, some linguists have focused on the fact that, in natural language, expressions that denote propositions can occur in sentence positions where expressions that appear to denote individuals can occur. The expressions in question are referred to as complement phrases (CP) and determiner phrases (DP), respectively, and the sentences that have positions where both CPs and DPs can occur may be called CP/DP-neutral constructions. Liefke (2014) and Liefke and Werning (2018) compile a wide variety of these and other similar constructions. Here are just a few examples:

1. DP/CP neutrality:
   a. Mary noticed [$_{\text{DP}}$ Bill].
   b. Mary noticed [$_{\text{CP}}$ that Bill waiting for Pat].

---

up just being *defined into* the modal operators, thereby making the principles definitional truths. In object theory, the principles in question aren't simply true by definition.

2. DP/CP coordinability:

   a. Mary remembered [$_{\text{DP}}$ Bill] and [$_{\text{CP}}$ that Bill was waiting for Pat].

3. CP nominalization:

   a. [$_{\text{DP}}$ Mary] bothered Bill.
   b. [$_{\text{CP}}$ That Pat was so evasive] bothered Bill.

4. DP/CP equatability:

   a. [$_{\text{DP}}$ The problem] was [$_{\text{DP}}$ Pat's dislike of Bill].
   b. [$_{\text{DP}}$ The problem] was [$_{\text{CP}}$ that Pat did not like Bill.]

5. Proposition-type anaphora:

   a. Mary told John [$_{\text{CP}}$ that it was raining]. John did not believe [$_{\text{PRO}}$ it].

Whereas Partee (2009) uses such constructions to question the distinction between the primitive types e and t in FTT, Liefke (2014) and Liefke and Werning (2018) conclude that such constructions (and others) provide evidence for developing a semantics for natural language based on single primitive type o, which is nevertheless to be interpreted as a higher Montagovian type $(s, (s, t))$. In Liefke (2014) (18, 86, 97, 163), this higher type is understood to be that of *propositional concepts*, i.e., functions from possible worlds to Montagovian propositions. By contrast, Liefke and Werning (2018) interpret the type $(s, (s, t))$ as the type for functions from contextually specified situations to sets of situations.[25] They then interpret both CPs and DPs in the higher type $(s, (s, t))$. But to give this analysis, they must introduce the notions of *situation*, *contextual specification of a situation*, and *situative proposition*, and invoke a rich ontology that includes worlds, times, locations, situations, inhabitants of situations, situative propositions, etc. For the most part, they stipulate the structure that is needed, e.g., a partial ordering $\sqsubseteq$ ('inclusion') on a set of situations, with top and bottom elements, etc.

In typed object theory, one can offer an alternative analysis of the linguistic data, namely, that the constructions involve nominalized propositions, i.e., abstract individuals that are defined by, and so correspond to, propositions. Typed object theory has a natural way to do this: for each proposition p, there is an abstract individual of type i that is the nominalization of p. Let p be any proposition, i.e., entity of type $\langle\,\rangle$, x be a variable ranging over individuals, $A!$ (*being abstract*) be a property of individuals and F a variable ranging over properties of individuals. Then object theory guarantees that the following definition picks out a canonical individual, $\check{p}$, which we may call *the nominalization of p*:

$$\check{p} \;=_{df}\; \iota x(A!x \;\&\; \forall F(xF \equiv F = [\lambda y\ p]))$$

[25] I shall continue to use parenthesis to denote derived, functional types in FTT, and use angled brackets to denote derived, relational types in RTT. But the reader should note that Liefke & Werning use angled brackets for Montagovian functional types.

This identifies the nominalization of p as the abstract individual that encodes just the property *being such that p* (i.e., encodes just $[\lambda y\ p]$). Given such a definition, we may interpret sentences like the ones above as giving rise to contexts in which the DPs and CPs both denote individuals. For example, *notice* can be a verb of type $\langle i, i\rangle$, so that (1.a) and (1.b) above can be analyzed, respectively, as follows, where W is the relation *waiting for*:

$$N(m, b)$$

$$N(m, \check{}[\lambda\ Wbp])$$

In the second case, *notice* relates Mary to the nominalization of the proposition *that Bill was waiting for Pat*. Thus, instead of type-raising, object theory *lowers* the relational type for propositions $\langle\ \rangle$ to the type for individuals!

This kind of solution then generalizes to the other cases, though one may have to apply certain operations on individuals, for example, to analyze the compound individuals such as the conjunction of the individual Bill and the nominalization of the proposition *that Bill was waiting for Pat* (to handle examples like 2.a).[26]

This ability to nominalize propositions in typed object theory is similar to its ability to nominalize properties. Suppose G is a property of individuals, i.e., of type $\langle i\rangle$. Then we may define the nominalization of G, written $\check{G}$, as follows:

$$\check{G} = \iota x(A!x\ \&\ \forall F(xF \equiv F = G))$$

In other words, the nominalization of G is the abstract object that encodes just G and no other properties. This allows the semanticist to give a uniform analysis of the sentences:

John is fun.
Fj
Running is fun.
$F\check{R}$

[26] Typed object theory provides such compound individuals. Where y denotes any individual and $\check{p}$ denotes the individual which is the nominalization of the proposition p, object theory asserts that there is an *intersect* object, $y \wedge \check{p}$, that encodes exactly the properties that y and $\check{p}$ exemplify in common:

$$y \wedge \check{p}\ =_{df}\ \iota x(A!x\ \&\ \forall F(xF \equiv Fy\ \&\ F\check{p}))$$

as well as a *union* object, $y \vee \check{p}$, that encodes all and only the properties exemplified by either y or $\check{p}$:

$$y \vee \check{p}\ =_{df}\ \iota x(A!x\ \&\ \forall F(xF \equiv Fy \vee F\check{p})).$$

In these representations, both ‘John’ and ‘Running’ denote individuals, though the latter is an abstract individual. Similarly, if G is a 3-place relation among individuals, e.g., x *gives* y to z, then we can identify its nominalization (*giving*) as the nominalization of the property that results by existentially projecting G to $[\lambda x\ \exists y \exists z Gxyz]$, i.e., as $^{\vee}[\lambda x\ \exists y \exists z Gxyz]$. So if *being rewarding* (R) is a property of individuals, we have the following analysis:

Giving is rewarding.

$R^{\vee}[\lambda x\ \exists y \exists z Gxyz]$

Note that in this analysis, no type-raising is involved.

Thus, where FTT systems often use type-raising (type-lifting) techniques to unify the analysis of natural language, typed object theory can analyze many constructions without such techniques. We’ve already seen some examples. Type-raising isn’t needed for the analysis of the intensions of natural language expressions, nor for the nominalizations of propositions and properties. Consider also the classic FTT analysis of using generalized quantifiers to unify the noun phrases ‘John’ and ‘every person’ by type-raising. Both expressions are often analyzed extensionally in FTT systems as denoting a set of properties of individuals, i.e., as $\{F \mid Fj\}$ and $\{F \mid \forall x(Px \rightarrow Fx)\}$, respectively. But in object theory, type-raising isn’t needed: both expressions have type i. ‘John’ denotes an individual and ‘every person’ can denote the following individual:

$$\imath x(A!x \;\&\; \forall F(xF \equiv \forall y(Py \rightarrow Fy)))$$

So ‘every person’ would denote the abstract object that encodes all the properties exemplified by every person.

Nor does type-raising help with fictions. In FTT systems, it is suggested that ‘Sherlock Holmes’ (‘h’) denotes $\{F \mid Fh\}$, i.e., a set of properties. But if that set is to be something other than the empty set, ‘h’ must have a denotation. Object theory provides such a denotation:

$$h = \imath x(A!x \;\&\; \forall F(xF \equiv CD \models Fh))$$

This identifies Holmes the abstract object that encodes exactly the properties F such that, in the Conan Doyle novels, Holmes exemplifies F.[27] Thus, Holmes is identified on the basis of the body of story-truths of the form: in the Conan Doyle novels, Holmes is F. By being abstract, Holmes is not a possibly concrete object. As Kripke noted, there are too many complete, possible objects consistent with the novels (supposing the novels are consistent). Holmes is an individual that is

[27] Note here that ‘In the Conan Doyle novels, p’ has been represented as a claim of the form ‘$s \models p$’ (‘p is true in s’), where s is a situation and p is a proposition. So truth in a situation is given the same analysis as truth at a world, namely, as $s[\lambda y\ p]$, i.e., s encodes the propositional property *being such that* p.

incomplete with respect to his encoded properties, but complete with respect to his exemplified properties. Given that the English copula is ambiguous between encoding and exemplification predication, we may say that Holmes 'is' a detective in the sense of *encodes*, but fails to exemplify detectivehood.

This analysis extends to fictional properties of individuals, such as *being a hobbit* (H). We don't need type-raising to interpret the predicate 'hobbit', for its analysis is similar to the analysis of names of fictional individuals: H denotes an abstract property of individuals, i.e., an abstract property with type $\langle i \rangle$. An abstract property of individuals encodes properties of properties of individuals. So where H is of type $\langle i \rangle$, x is a variable of type $\langle i \rangle$, and $A!$ is a constant and F a variable of type $\langle\langle i \rangle\rangle$, we may identify *being a hobbit* as follows:

$$\textsc{h} = \imath x(A!x \;\&\; \forall F(xF \equiv \mathit{LordOfTheRings} \models F\textsc{h}))$$

That is, *being a hobbit* is the abstract property of individuals that encodes exactly the properties of properties of individuals that *being a hobbit* exemplifies in *The Lord of the Rings*.[28]

In summary, then, typed object theory avoids type-lifting by taking advantage of the abstract objects that exist at each type. It is based on RTT with a single primitive type and offers a natural way to define situations, possible worlds, fictional entities, etc. These entities have precise definitions and the main principles governing them can be derived. Propositions and properties both, no matter whether simple or complex, have nominalizations, and we need not interpret sentence positions that are neutral with respect to CPs and DPs as positions requiring a higher type.

4.4 Intensional Type Theory: II

We now turn to a discussion that compares typed object theory to the intensional framework developed in Williamson (2013). We begin by showing that an argument Williamson raises against general (Henkin) models can be undermined.

4.4.1 Response to an Argument Against Henkin Models

In 2013 (226–230), Williamson develops an extended argument that is designed to show the superiority of *standard* models of higher-order logic to *general* (Henkin) models. In this argument he distinguishes the standard notions of logical

[28]Strictly speaking, in this analysis of 'hobbit' and in the analysis of the name 'Holmes', we need to index the term being analyzed to the story in question. So we should use H_{LOTR} and h_{CD} on both sides of the identity symbol in the respective principles.

consequence and validity from the analogous notions, *g-logical consequence* and *g-validity*, that apply to general models. Williamson begins by asserting (2013, 226):

> Despite the formal tractability of g-logical consequence, general models are more complex and less natural than standard models. Why have arbitrary restrictions on the permissible intensions of the appropriate type for a predicate?

Though Williamson goes on to give an example, the second sentence in this opening statement betrays a presupposition that is rejected by typed object theory, namely, that the *permissible* intensions are those that are given by possible world semantics, in which relations are identified as set-theoretically defined functions from worlds to sets of n-tuples. This presupposes that set theory and set membership offer a more fundamental account of relations than a direct, axiomatic theory of relations and predication. But object theory has no such presupposition. From the point of view of typed object theory, we should reverse the order: the permissible intensions are those that are given by a mathematically precise theory of relations, such as the one offered by (typed) object theory. If there is nothing more fundamental than individuals, relations, and predication, why suppose that set theory with possible worlds as urelements gives us a greater insight as to what relations or intensions exist?

Once we recognize this presupposition in Williamson's argument, it becomes easy to undermine the other reasons Williamson gives for preferring standard models to general models. He notes, for example, that the Comprehension Principle is standardly valid, but not g-valid (2013, 228):

> By contrast, in some general models, dom($\langle t_1, \ldots, t_n \rangle$) omits the intensions of $\lambda v_1 \ldots v_n(A)$ needed as a value of V to verify an instance of Comp, so Comp is not g-valid.

But this offers no reason why standard models should be preferred. Why let the semantics drive a precise theory of relations? Instead, the comprehension principle for relations, which is derivable in object theory (suitably restricted to exclude encoding subformulas from allowable matrices), should drive the semantics. This comprehension principle tells us the conditions under which *relations* exist. I suspect that the reason Williamson doesn't consider it definitive is the same one we encountered before in trying to understand why FTT rather than RTT became standard in linguistics: without encoding formulas to give precise identity conditions for relations, Williamson has no theory of relation identity to fall back on, other than the set-theoretic reconstruction of relations as Montagovian intensions. But, from the point of view of typed object theory, a general model is sufficient if it makes the comprehension principle for relations valid. Such general models would then include everything needed to show that the theory of relations is consistent.

Williamson discusses this option (2013, 229):

> We could add Comp as an extra principle to Gallin's axiomatic system presented earlier, and restrict the general models to those in which it is valid. Of course, the resulting logic would still have a recursively axiomatizable set of theorems, and so be weaker than the standard logic. Even a general model that validates Comp may have highly restricted intensions for most types because many intensions correspond to no formula of the language, relative to any values of its parameters.

But again, Williamson assumes that general models would have ‘highly restricted intensions’, because he supposes that the *intensions* are given by set-theoretic functions from possible worlds to sets of n-tuples. We should not, however, accept such a prior characterization of intensions. That is to give the conception of intensions derived from set-theory preference over the conception of intensions derived from metaphysical considerations. Object theory starts with a primitive, fine-grained notion of relations; these are *more fine-grained* than set-theoretic functions from worlds to sets of n-tuples: we saw in Sects. 4.1.1 and 4.1.2 that while Montagovian intensions collapse necessarily equivalent relations, the identity conditions for relations in typed object theory do not.

Williamson next charges that the structure of standard models, but not g-models, is what our metaphysics *should* characterize (2013, 229–230):

> Thus non-standard models also differ from standard ones in respects relevant to the evaluation of claims about purely logical structure, in the sense of claims expressed by formulas without non-logical constants. But logical structure is what the logical core of our metaphysics is supposed to characterize. ... Hence a g-logic is less informative than standard logic about purely logical structure. A metaphysical theory based on g-logic rather than standard logic is neutral on many of the very questions it is supposed to answer.

But I would reply that it is exactly the neutrality of g-models that prevents it from falling into the obvious error of standard models, namely, the error of collapsing relations, properties, and propositions that are necessarily equivalent. A g-logic *should* remain neutral on many questions that should be decided on the basis of theory, not on the basis of the *set-theoretic artifacts* of a standard model.

Finally, Williamson claims (2013, 230):

> Moreover, to the extent to which we take models for $\mathtt{ML_P}$ seriously, the standard ones are more faithful than the non-standard ones to our intended interpretation.

This strikes me as rather controversial. Given the precise theory of relations offered in (typed) object theory, how could standard models based on Montagovian intensions, which collapse relations that can be kept apart in g-models of object theory, be more faithful?

4.4.2 Comparison of the Ontologies

If we put his argument against g-models aside, though, there are some interesting points of comparison between the typed object theory and Williamson’s intensional logic. One is that typed object theory uses one set of types for both its syntax and semantics. Williamson, by contrast, uses one set of types for the syntax of his language and a different set of types for its semantics. For the syntax, he uses the standard RTT types, though using e as the primitive type for individuals. For the semantics, he adds w as a second base type, and then defines a new type $\tau\sigma$ from each σ in the syntactic hierarchy, as follows (2013, 236–7):

> Each type t of $\texttt{ML}_\texttt{P}$ corresponds to a type τt of the metalanguage by the rule τe and $\tau\langle t_1, \ldots, t_n\rangle$ is $\langle \tau t_1, \ldots, \tau t_n, w\rangle$. But we add a cumulative infinite limit type λ to the metalanguage: the expressions of type λ are exactly those of any finite type. Thus expressions of type λ belong to some more specific type, but expressions of type $\langle\lambda\rangle$ do not.

If I'm understanding this correctly, then this typing scheme, unlike that of typed object theory, essentially takes the entities denoted by n-place relational predicates to be $n + 1$-place relations and requires the metaphysician to regard relations essentially as world-indexed entities. This fundamentally changes the way in which relations are to be conceived and such a change is not required by typed object theory. Williamson would no doubt justify the proposal by citing the advantages of the semantics he goes on to give (237–8) (i.e., a kind of homophonic semantics in which quantification can be conceived without domains and as unrestricted). But that semantics, as Williamson admits, requires a *plural* conception of higher-order quantifiers, something that doesn't easily generalize to relations, given that there seems to be no natural way to render quantification over relations in the plural idiom.

Moreover, Williamson's world-indexed relations leave open a variety of questions. If F and G are variables for properties with type $\langle\sigma\rangle$ and x a variable for an object of type σ, do the world-indexed relations obey the law: $\forall x \forall w(Fxw \equiv Gxw) \rightarrow F = G$? What is the denotation of complex λ-expressions in the semantics that Williamson develops? His semantics (2013, 238) doesn't say.

By contrast, typed object theory simply rests with its axiomatic foundations; there are axioms for quantification, axioms for relations, and definitions and theorems governing possible worlds. No set-theoretic model of such a system gives any deeper insight into the nature of the entities being described. One should not mistake the entities in such models or the artifactual set-theoretic domains of the models for the entities and notions being described.

4.5 Conclusion

It might be thought that RTT, despite its elegance in having a single primitive type, can be reduced to FTT. But Oppenheimer and Zalta (2011) show that FTT has no straightforward way of representing the logic of typed object theory as the latter is formulated in RTT. This suggests that RTT is the more general framework. Basically, we noted that in FTT, every formula can be converted into a term. The semantics of the quantified formula $\forall x^\sigma \varphi$ is handled by converting φ to $[\lambda x^\sigma \varphi]$, which is a function that maps objects of type σ to a truth value. Then $\forall$ is interpreted as a particular function that maps the expression $[\lambda x^\sigma \varphi]$ to a truth value. In particular, $\forall$ maps $[\lambda x^\sigma \varphi]$ to The True, i.e., $\forall x^\sigma \varphi$ is true, just in case the function $[\lambda x^\sigma \varphi]$ maps every object to The True.

But in typed object theory, formulas with encoding subformulas *can't* be converted to terms, on pain of paradox. The formula $xF \,\&\, \neg Fx$ can't be converted to $[\lambda x\; xF \,\&\, \neg Fx]$ since the latter is not even well-formed in object theory.

Oppenheimer and Zalta (2011) point out that FTT can't therefore interpret the expression $\forall F(xF \,\&\, \neg Fx)$ by applying the higher order function $\forall$ to the predicate $[\lambda x\ (xF \,\&\, \neg Fx)]$, since the latter isn't in the language.

These considerations, as well as the ones presented earlier in the body of this paper, may prove helpful when comparing the relative merits of FTT and RTT systems for the analysis of natural language, and when comparing foundations that take relations and predication as basic, instead of functions and function application or sets and set membership.

Bibliography

Anderson, C. A. (1989). Russellian intensional logic. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 67–103). New York: Oxford University Press.

Anderson, D. J., & Zalta, E. (2004). Frege, boolos, and logical objects. *Journal of Philosophical Logic, 33*(1), 1–26.

Boolos, G. (1987). The Consistency of Frege's *Foundations of Arithmetic*. In J. Thomson (Ed.), *On being and saying*. Cambridge: MIT Press; reprinted in Boolos, G. (1998) J. Burgess & R. Jeffrey (Eds.), *Logic, logic, and logic* (pp. 183–201). Cambridge: Harvard University Press.

Bueno, O., Menzel, C., & E. Zalta, (2014). Worlds and propositions set free. *Erkenntnis, 79*, 797–820.

Carnap, R. (1929). *Abriss der Logistik*. Wien: J. Springer.

Church, A. (1940). A formulation of the simple theory of types. *The Journal of Symbolic Logic, 5*(2), 56–68.

Church, A. (1974). Russellian simple type theory. *Proceedings and Addresses of the American Philosophical Association, 47*, 21–33.

Cresswell, M. (1975). Hyperintensional logic. *Studia Logica, 34*, 25–38.

Cresswell, M. (1985). *Structured meanings: The semantics of propositional attitudes*. Cambridge: Bradford Books/MIT Press.

Fox, C., & Lappin, S. (2001). A Framework for the hyperintensional semantics of natural language with two implementations. In P. de Groote, G. Morrill, & C. Restoré (Eds.), *LACL 2001: Proceedings of the 4th International Conference on Logical Aspects of Computational Linguistics, Le Croisic, June 27–29, 2001* (Lecture notes in artificial intelligence 2099. pp. 175–92). Berlin: Springer.

Gallin, D. (1975). *Intensional and higher-order modal logic: With applications to Montague semantics*. Amsterdam: North-Holland.

Kripke, S. (1973). [2013] *Reference and existence: The John Locke lectures*. Oxford: Oxford University Press.

Liefke, K. (2014). *A single-type semantics for natural language*. Enschede: Ipskamp Drukkers.

Liefke, K., & Werning, M. (2018). Evidence for single-type semantics – an alternative to *e*/*t*-based dual-type semantics. *Journal of Semantics, 35*(4), 639–685.

Linsky, B., & Zalta, E. (1995). Naturalized platonism vs. platonized naturalism. *The Journal of Philosophy, 92*(10), 525–555.

Meinwald, C. (1992). Good-bye to the third man. In: R. Kraut (Ed.), *The Cambridge companion to Plato* (pp. 365–396). Cambridge: Cambridge University Press.

Menzel, C., & Zalta, E. (2014). The fundamental theorem of world theory. *Journal of Philosophical Logic, 43*(2), 333–363.

Montague, R. (1973). The proper treatment of quantification in ordinary english. In J. Hintikka, J. Moravcsik, & P. Suppes (Eds.), *Approaches to Natural Language: Proceedings of the 1970 Stanford Workshop on Grammar and Semantics* (pp. 221–242). Dordrecht: D. Reidel. reprinted

in R. Thomason (Ed.). (1974). *Formal philosophy: Selected papers of richard Montague* (pp. 247–270). New Haven: Yale University Press, [Page reference is to the reprint.]
Muskens, R. (1989). A relational formulation of the theory of types. *Linguistics and Philosophy, 12*, 325–346.
Muskens, R. (1995). *Meaning and partiality*. Stanford: CSLI Publications.
Muskens, R. (2007). Intensional models for the theory of types. *The Journal of Symbolic Logic, 72*(1), 98–118.
Nodelman, U., & Zalta, E. (2014). Foundations for mathematical structuralism. *Mind, 123*(489), 39–78.
Oppenheimer, P., & Zalta, E. (2011). Relations versus functions at the foundations of logic: Type-theoretic considerations. *Journal of Logic and Computation, 21*: 351–374.
Orey, S. (1959). Model theory for the higher order predicate calculus. *Transactions of the American Mathematical Society, 92*, 72–84.
Partee, B. (2009). [2007], Do we need two basic types? *Snippets*, Issue 20 (Special issue in S. Beck & H.-M. Gärtner (Eds.), *Honor of Manfred Krifka*) (pp. 37–41). Milan: Edizioni Universitarie di Lettere Economia Diritto; see also Type Theory and Natural Language: Do We Need Two Basic Types? 2007 presentation handout, 100th Meeting of the Seminar: Mathematical Methods Applied to Linguistics, 31 Mar, Moscow State University. http://people.umass.edu/partee/docs/TwoTypesHOMarch07.pdf.
Pelletier, F.J., & Zalta, E. (2000). How to say goodbye to the third man. *Noûs, 34*(2), 165–202.
Pollard, C. (2005). Hyperintensional semantics in a higher-order logic with definable subtypes. In M. Fernández, C. Fox, & S. Lappin (Eds.), *Lambda calculus, type theory, and natural language* (pp. 32–46). https://pdfs.semanticscholar.org/2562/1a82b55c0d34f90c6b15ad83939c0e04ec1c.pdf.
Pollard, C. (2008). Hyperintensions. *Journal of Logic and Computation, 18*, 257–82.
Quine, W. V. O. (1960). Variables explained away. *American Philosophical Society, 104*(3), 343–347; reprinted in Quine, W. V. O. (1966). *Selected logical papers* (pp. 227–235). New York: Random House.
Russell, B. (1908). Mathematical logic as based on the theory of types. *American Journal of Mathematics, 30*(3), 222–262.
Schütte, K. (1960). Syntactical and semantical properties of simple type theory. *The Journal of Symbolic Logic, 25*(4), 305–326.
Thomason, R. (1980). A model theory for propositional attitudes. *Linguistics & Philosophy, 4*, 47–70.
van Benthem, J., & Doets, K. (1983). Higher-order logic. In D. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic* (Vol. I, pp. 275–329). Dordrecht: D. Reidel.
Williamson, T. (2013). *Modal logic as metaphysics*. Oxford: Oxford University Press.
Zalta, E. (1982). Meinongian type theory and its applications. *Studia Logica, 41*(2–3), 297–307.
Zalta, E. (1983). *Abstract objects: An introduction to axiomatic metaphysics*. Dordrecht: D. Reidel.
Zalta, E. (1988a) *Intensional logic and the metaphysics of intentionality*. Cambridge: MIT Press.
Zalta, E. (1988b) A comparison of two intensional logics. *Linguistics and Philosophy, 11*(1), 59–89.
Zalta, E. (1993). Twenty-five basic theorems in situation and world theory. *Journal of Philosophical Logic, 22*, 385–428.
Zalta, E. (1997). A classically-based theory of impossible worlds. *Notre Dame Journal of Formal Logic, 38*(4), 640–660.
Zalta, E. (1999). Natural numbers and natural cardinals as abstract objects: A partial reconstruction of Frege's *Grundgesetze* in object theory. *Journal of Philosophical Logic, 28*(6), 619–660.
Zalta, E. (2000a) Neo-Logicism? An ontological reduction of mathematics to Metaphysics. *Erkenntnis, 53*(1–2), 219–265.
Zalta, E. (2000b) The road between pretense theory and object theory. In A. Everett & T. Hofweber (Eds.), *Empty names, fiction, and the puzzles of non-existence* (pp. 117–147). Stanford: CSLI Publications.

Zalta, E. (2000c) A (Leibnizian) theory of concepts. *Philosophiegeschichte und logische Analyse/Logical Analysis and History of Philosophy, 3*, 137–183.

Zalta, E. (2001). Fregean senses, modes of presentation, and concepts. *Philosophical Perspectives* (*Noûs* Supplement)*, 15*, 335–359.

Zalta, E. (2006a) Essence and modality. *Mind, 115*(459), 659–693.

Zalta, E. (2006b) Deriving and validating Kripkean Claims using the theory of abstract objects. *Noûs, 40*(4), 591–622.

Part II
Fictionalism or Realism in Philosophy of Mathematics

Chapter 5
Contingent Abstract Objects

Otávio Bueno

Abstract It is usual to claim that mathematical objects, which are typically thought of as abstract entities, exist necessarily and that mathematical truths are necessary. In this paper, I resist this view, arguing instead that mathematical objects are contingent and that statements about them are not necessarily true (if true at all). I provide an account of the source of the apparent necessity of mathematics, and argue that, despite its ubiquity, nothing requires the acceptance of this received view. As an alternative, I offer an alternative, non-necessitarian conception of abstract objects, which recognizes the contingency of mathematical objects.

Keywords Mathematical objects · Contingency · Mathematical necessity · Platonism · Agnosticism · Mathematical pluralism

5.1 Introduction

According to the received view of mathematical objects, if they exist, they necessarily exist, and statements about them, if true, are necessarily true. This provides a particular conception of abstract objects that supports platonism, a view that asserts the antecedents of both conditionals above. It insists that abstract objects (and, in particular, mathematical entities) exist and statements about them have truth-value. In this paper, I resist the received view, and argue instead that mathematical objects do not necessarily exist nor are statements about them necessarily true, if true at all. I offer an account of the apparent necessity of the relevant mathematical statements and why mathematical objects seem to exist necessarily. Once the source of the apparent necessity is understood, it becomes clear that, despite the ubiquity of the received view, nothing requires its acceptance. In fact, I argue, there are arguments to the effect that an alternative, non-necessitarian conception of abstract objects, which acknowledges the contingency of mathematical entities, should be taken seriously.

O. Bueno (✉)
Department of Philosophy, University of Miami, Coral Gables, FL, USA

J. L. Falguera, C. Martínez-Vidal (eds.), *Abstract Objects*, Synthese Library 422,
https://doi.org/10.1007/978-3-030-38242-1_5

5.2 The Traditional View: Platonism

5.2.1 Platonism's Central Features

Platonism is a view about the ontology of mathematics, which is committed to the existence of mathematical objects, or mathematical structures, properties, and relations, depending on the particular versions of the view. For simplicity, I will talk about objects, but what I have to say extends, without difficulty, to mathematical properties, relations and structures as well. According to this view:

(P1) Mathematical objects (structures, properties, relations) exist.
(P2) Mathematical objects are abstract, that is, they are non-spatiotemporal and are causally inert.
(P3) Mathematical objects are independent of what is said or thought about them.
(P4) Mathematical statements, if true, are necessarily true.

This view combines metaphysical and semantic assumptions. Metaphysically, platonism involves claims about the existence of mathematical objects (P1); about the nature of these objects, highlighting their abstract character (P2), and about the independence of the existence of mathematical objects and one's processes of describing or thinking about them (P3). Semantically, platonism involves a particularly strong form of realism according to which not only mathematical statements have truth-values but also, if true, they are necessarily so (P4).

There are different forms of platonism. On the one hand, standard platonism emphasizes the importance of mathematical objects for the proper understanding and characterization of mathematics (Gödel 1964): mathematics is ultimately about these objects (as well as the relations among them, the relations among these relations, and so on). These are the objects that exist, albeit not in space-time, independently of those who think or talk about them. On the other hand, structuralist platonism stresses the role of mathematical structures, rather than objects, as the basic component of the ontology of mathematics (Resnik 1997; Shapiro 1997). These views typically question whether there is anything more to mathematical objects than the role they play in particular structures. As a result, the emphasis lies on structures rather than objects, which are then taken to be nothing more than placeholders in the structures under consideration. On this view, there is nothing more to the number 1, in the structure of natural numbers, than being the successor of 0. What matters is its place in the overall number-theoretic structure. For the structuralist platonist, structures are the proper locus of ontological commitment, independently of how they are described or thought.

There is much to be said about these families of views. I will focus on three aspects, which will be important in what follows to clarify some salient traits of platonism, namely, (i) existence (particularly what, in mathematics, is taken for mathematical objects to exist, in contrast with requirements often articulated in metaphysics); (ii) necessity (and, in particular, three ways of understanding it), and (iii) mathematical universes (and some of their general features).

5.2.2 Existence

What does it take for mathematical objects to exist? It is important to consider, first, what it takes, within mathematical practice, for the relevant objects to exist. As an illustration, I will focus on a typical example. In topology, it is the case that to every metric space there is a topological space. The claim clearly states the existence of a topological space given a metric space. As formulated here, existence is understood as what can be established on the basis of the relevant principles, namely, those that characterize the mathematical objects under consideration. In this example, the existence of a topological space follows from the assumption that there is a metric space. How is this established?

One starts by formulating the relevant concepts. A metric space is a set for which the distance between all members of the set is defined. A topology is a collection of subsets of a nonempty set, called *open sets*, such that (i) the empty set and the entire set are open; (ii) the union of an arbitrary collection of open sets is open; and (iii) the intersection of a finite number of open sets is open. How is a topological space obtained from a metric space? Given a metric space M, one obtains a topology on M by specifying that a set T is open just in case for every x in T, there exists a positive number e such that the open balls $B_e(x)$ of radius e centered around x are constructed. It is then established that the open balls in question satisfy the conditions of a topology. As this example illustrates, existence in mathematics involves obtaining objects (in this case, open sets that form a topology) *via* suitable mathematical constructions (such as the obtaining of open balls) from particular mathematical assumptions (in this instance, the existence of a metric space). In mathematics, existence is then connected to such construction processes, which are typically part of the inferential practices involved in proving mathematical theorems.

In contrast, a metaphysical conception of existence involves adding a substantive interpretation to the content of what is taken to exist. For instance, some argue that existence is ontological independence of one's linguistic practices and psychological processes (Azzouni 2004). On this conception, those things that are ontologically independent from the ways in which they are described or are thought of exist. Given that, according to platonism, mathematical objects do *not* depend on how one may describe or think about them, it follows from this conception of existence, that mathematical entities exist. In fact, these objects would have existed even if no one has ever thought about or described them. Interestingly, this is not the conclusion Azzouni (2004) draws from the ontological independence conception of existence. He defends a form of nominalism according to which, mathematical objects do *not* exist, given that they are entirely made up by us, and thus are ontologically *dependent* on our linguistic practices and psychological processes. As a result, ontological independence ends up supporting conflicting views about the ontology of mathematics, relative to whether one takes mathematical objects to be created by mathematicians, and thus be dependent on traits of mathematical activity, or be discovered by them, and thus be independent of how these objects are constituted in mathematical practice.

The notion of ontological independence highlights the distinction between what exists and what is dependent on human discourse and thought. It is a metaphysical distinction, separating categories of different kinds: the existent and the human-dependent. Not surprisingly perhaps, this notion of existence goes well beyond anything that is stated and explicitly found in mathematical practice and in the formulation of mathematical results. Returning to the example from topology above, nothing in the characterization of metric space, topological space, open set or open ball requires taking any stand on whether the objects in questions are ontologically independent of one's linguistic practices or psychological processes. Talk of ontological independence provides an *additional* layer of interpretation that is appended to the content of the mathematical result that associates the existence of a topological space to each metric space. The result goes through quite separately from the metaphysical issue of how the existence of such a topological space is understood (whether in terms of ontological independence from one's linguistic practices and psychological processes or in some other way).

It may be argued that this is not so. Mathematicians formulate the relevant concepts (of *metric space*, *topological space*, and *open ball*) and whether a given theorem holds or not ultimately depends on these concepts. Thus, in the end, the results are indeed dependent on the linguistic practices and psychological processes that are central to mathematical activity, an activity that is crucially based on the relevant concepts.

There is, however, an equivocation here between ontological and conceptual dependence. The theorem to the effect that for each metric space there is a topological space depends *conceptually* on the characterization of the concepts of *topological* and *metric spaces*. One could not state, let alone prove, the theorem without invoking the relevant concepts. Arguably these concepts are also created by mathematicians, who thought of, designed, and articulated them. (Those who think that concepts are abstract objects may demur, but such an account of concepts *prima facie* faces at least the same problems that platonism does.) Despite that, it does *not* follow that the existence of a topological space from a given metric space is *ontologically* dependent on the concepts of *metric* and *topological spaces*. The result formulated in the theorem depends *conceptually* on the relevant notions but *no* ontological dependence is involved between the concepts and the relevant objects. In fact, the objects in question play no role whatsoever, given that even if mathematical objects did not exist at all, mathematical practice would remain the same. Clearly, mathematical objects do not play any causal role in mathematical practice, given that such objects are causally inactive. Nor do mathematical objects play an epistemic role, given that theorems are formulated and proved without any access to them. Rather concepts are doing the relevant work, since it is in terms of them that mathematical statements are expressed and established. Mathematical theorems basically hold due to relations among concepts, to which mathematicians have clear access; relations among mathematical objects (if any) are irrelevant for this, since epistemic access is restricted to concepts, with objects dropping out of the picture. In the end, even if there were ontological dependences among mathematical objects, they would not be relevant to mathematical practice. (For a detailed argument to the effect that mathematical objects play no role in mathematical practice, see Azzouni 1994.)

In the end, as the example from topology illustrates, mathematical activity does not demand commitment to any such metaphysical conception of existence. The practice can be articulated and understood without the assumption that existence is the sort of ontological independence just discussed. Rather, a metaphysically deflationary understanding of existence in terms of what is established by suitable constructions is enough and reflects better what takes place in mathematical practice.

It may be argued that these considerations can be resisted by invoking an account of mathematical practice developed by Imre Lakatos in *Proofs and Refutations* (Lakatos 1976; I owe this comment to an anonymous reviewer). On this account, the argument goes, changes in a concept—that of a *polyhedron*—emerge in non-arbitrary ways and were guided by grasping something the concept was supposed to capture. Lakatos' insightful narrative tracks the changes made to the concept of *polyhedron* in the process of attempting to prove Euler's conjecture that states that the number of vertices (V), edges (E), and faces (F) of a regular polyhedron satisfies the equation: $V - E + F = 2$. The dynamics of attempted proofs, counter-examples, hidden lemmas, proof analysis, refined conjectures, and new proofs highlights the way in which the process of proving a result is basically a process of conceptual refinement of naïve conjectures.

Interestingly, in response to the objection, Lakatos' analysis reinforces rather than undermines the point that the existence of mathematical objects plays no role in mathematical discovery. After all, nowhere in the process of establishing a naïve conjecture has the grasping of objects (that the concepts invoked were supposed to capture) any role. The process is internal to the concepts in question. The actual objects play no part in the process. Nothing would have changed if none of the objects under consideration existed at all.

5.2.3 Necessity

With regard to the notion of necessity in mathematics, three conceptions should be considered: (a) necessity in terms of possible worlds, (b) necessity as invariability, and (c) necessity as unconditional truth. I will consider each of them in turn.

(a) *Necessity as truth in all possible worlds*: This is a modal realist conception of necessity in general and of mathematical necessity in particular. Its most thorough defense can be found in David Lewis' work (see, especially, Lewis 1986). According to modal realism, there is a plurality of worlds, each of which is roughly maximally connected spatiotemporal regions. On this conception, the modal status of a proposition is tied to its truth in a possible world: necessary ones are true in all worlds; possible ones are true in some. By quantifying over worlds, the modal realist aims to avoid commitment to any primitive modal notion, given that worlds, at least according to Lewis, do not involve any such modality. However, it is unclear that this goes through: the logical space (of all worlds) need to include all and only the worlds that are possible, but that is not a possibility that can be itself formulated in terms of worlds (see Shalkowski 1994 as well as Bueno and Shalkowski 2015 for a critique).

In principle, mathematical statements if true are necessarily true (they are true in all worlds). After all, there is nothing about worlds and their spatiotemporal traits that undermine truths that depend neither on the concreteness of such worlds nor on any such traits. It then follows that there is no contingency in abstract objects: on this conception, their properties, whatever they ultimately are, hold necessarily.

(b) *Necessity as invariability*: A different conception of necessity articulates the content of necessary truths in terms of invariability: if something is necessary, it does not change across possibilities. On this conception, "[a] truth is necessary if it describes matters that don't vary across [. . .] possible alternatives" (Kment 2014, p. 20). The necessary is then the invariant: alternatives change, but what is necessary remains the same. This conception of necessity captures and highlights a significant aspect of the necessary: its invariance. This is a feature that the modal realist can, of course, accommodate in terms of possible worlds. What is true in all possible worlds is invariant across all possible alternatives. Possible worlds offer a framework to articulate the invariability conception, but one need not assume such worlds to characterize necessity as invariability, since worlds offer just one way to express and formulate possible alternatives. This conception specifies that it is the invariability that makes the necessary (rather than the other way around), however it ends up being formulated (in terms of worlds, states of affairs, abstract objects, or something else altogether).

On the traditional platonist conception, since mathematical objects do not vary at all across possibilities (being abstract, they are neither spatiotemporally located nor causally active, and thus are immutable), all truths about them (all mathematical truths) are indeed necessary. As a result, the connection between the traditional conception and the understanding of necessity in terms of invariability can be established: a significant feature of the former is captured by the latter.

(c) *Necessity as unconditional truth*: According to this conception, necessities obtain and are what they are independently of whatever turns out to be the case. "The metaphysical necessities are those propositions that hold unconditionally—i.e., independently of what else is the case. By contrast, a true proposition is a *contingent* truth if its truth is contingent on something else, i.e. dependent on what else is the case" (Kment 2014, p. 20). This conception captures a salient feature of necessity: its unconditional nature.

Modal realists can also accommodate this conception of necessity. Given that, on their view, what is necessary is what is true in all models, necessary truths are unconditional: they do not depend on any particular feature of what else goes on in any world. Nothing that happens in any world undermines (or could undermine) any given necessary truth. Hence, the unconditional account goes through. Moreover, on the traditional platonist conception, mathematical truths are unconditional truths as well: they are truths that do not depend on anything else that is the case. Mathematical truths, on this view, are what they are independently of whatever goes on in the world. As a result, in this case, the traditional conception and the unconditionality account of necessity go hand in hand.

5.2.4 Mathematical Universes

On the traditional platonist conception, mathematics is about a domain of independently existing mathematical objects (properties, relations, and structures), which, taken together, form a universe (or universes). The understanding of what such universes are like depends on whether platonism is characterized as being committed to the uniqueness of the mathematical universe (a Gödelian conception; see Gödel 1964) or to the multiplicity of mathematical universes (full-blooded platonism; see Balaguer 1998).

According to the uniqueness view, there is just one mathematical universe. Mathematical statements are true provided they describe correctly actual mathematical objects, their properties, relations, and the structure they generate. Intended interpretations of mathematical theories (such as arithmetic or analysis) are true relative to this universe. The uniqueness of the mathematical universe also suggests that there is only one suitable interpretation of mathematical theories. If multiple non-equivalent interpretations are offered (such as when one is faced with an infinitude of non-standard models of analysis, arithmetic or set theory), then only one of them is adequate, namely, the one that properly captures the suitable structure of the mathematical universe. The remaining interpretations fail to do that in one respect or another.

In contrast, according to the multiplicity view, there is a plurality of mathematical universes. Mathematical statements are true just in case they describe correctly some part of the mathematical universes, including some of the actual mathematical objects, their properties, relations, and the structures they generate; but there is no single mathematical universe that uniquely exemplifies all such objects, properties, and relations. A multiplicity of mathematical universes is in place so that to every way a mathematical description (such as a mathematical statement or a theory) could be, there are mathematical objects (properties and structures) that are that way; in other words, every possible mathematical description is true of some portions of the mathematical universes. What are possible ways for mathematical descriptions to be? On classical formulations of the multiplicity view, in which the underlying logic is classical (Balaguer 1998), the possibility of a mathematical description is tied to its consistency. Thus, on this view, every consistent mathematical theory is true of some part of the mathematical universes. On non-classical formulations of the multiplicity view, in which the underlying logic is paraconsistent (Beall 1999), the possibility of a mathematical description is tied to its non-triviality (that is, not everything follows from a contradiction of the form *A* and not-*A*). Thus, on this view, every non-trivial mathematical theory is true of some part of the mathematical universes, even if that part turns out to be inconsistent. What is resisted by the non-classical formulation is the triviality of mathematical theories in the presence of inconsistency.

The multiplicity view, particularly in its classical version (Balaguer 1998), still insists that there is an intended conception of mathematical theories (such as arithmetic or analysis) that is meant to be captured by the intended models. So, despite the multiple theories that are allowed for in this framework, only one interpretation of these theories is ultimately adequate. Thus, a form of uniqueness is still found even within the multiplicity view.

Whether mathematical universes are thought of as being unique or not, mathematical objects, properties, relations, and structures are taken to exist independently of the way in which they are characterized and described. They are taken to exist necessarily. After all, on the traditional formulation, true mathematical statements are true in all possible worlds: given that worlds are concrete (that is, as noted, spatiotemporally located and causally active), they do not contribute to the specification of the relevant mathematical content, which is abstract (and, thus, completely independent of any spatiotemporal or causal constraints).

Alternatively, if worlds are not to be invoked in the formulation of necessity, mathematical statements also turn out to be necessary on both conceptions of the necessary discussed above. On the invariability account, true mathematical statements "describe matters that do not vary across [. . .] possible alternatives" (Kment 2014, p. 20), given that concrete possible alternatives do not affect abstract mathematical configurations. On the unconditional truth account, true mathematical statements hold independently of "what else is the case" (Kment 2014, p. 20), given that what is concretely the case does not affect what holds abstractly. Whether formulated in terms of possible worlds or not, on the platonist conception, mathematics is ultimately necessary.

Platonism about mathematics has the advantage of being able to accommodate the objectivity of mathematics without trouble. Given that, on this view, the existence of mathematical objects (properties, relations, and structures) is independent from what is said or thought about them, and given that such objects, in turn, are what they are independently of the way in which they are conceptualized or described, mathematics is indeed objective. On this view, it is the independent existence of the relevant objects that secure the objectivity of mathematics, in a perfectly analogous way to what happens in a realist approach to the sciences, where the objectivity of scientific theories results from the existence of the relevant objects. The truth of mathematical theories does not depend on what is thought about numbers any more than the truth of geological theories depend on what is thought about mountains. In both cases, the truths in question depend on independently existing mathematical and geological entities, respectively.

It is not my goal here to revisit the well-known challenges to platonism (for a survey, see, for instance, Bueno 2011). Rather I want to articulate an alternative conception of the ontology of mathematics in which the necessitarian and existential features of platonism are not required, thus making room for contingent abstract objects, but without commitment to their existence, and in such a way that the resulting account is still able to accommodate the objectivity of mathematics. I will turn to this conception now.

5.3 An Alternative Approach: Agnosticism

It is unclear that mathematical practice requires a commitment to platonism. Platonism provides a particular interpretation of mathematical practice, which adds to mathematical discourse a specific layer of metaphysics: a certain conception of the existence of mathematical objects and their nature, as abstract entities that exist independently of how they are formulated and described. But this conception is not forced upon us by mathematical practice. In fact, the practice is silent about the existence and independence of mathematical objects, understood in this metaphysical way, as the example of topology and metric space mentioned above illustrates. It is settled in the practice that to a metric space there is a topological space, but there is no commitment in the practice to the ontologically independent existence (or not) of such spaces. This is simply an issue the practice is silent about.

As an alternative, one can adopt a form of *nominalism*, according to which:

(N1) Mathematical objects (structures, properties, relations) do not exist.
(N2) Mathematical objects would be abstract if they existed (in this case, they would be non-spatiotemporal, causally inert entities).
(N3) Mathematical objects are not independent of one's linguistic practices and psychological processes.
(N4) Mathematical statements if true are not necessarily true.

This is, of course, the denial of the corresponding platonist claims (P1)–(P4) discussed above. Let me consider each of them in turn.

With (N1), the existence of mathematical objects is resisted. This can be done by adopting the same understanding of existence as the one held by platonists: ontological independence from one's linguistic practices and psychological processes (Azzouni 2004). But, as opposed to platonists, who insist that mathematical objects exist independently of the way in which they are thought of or characterized, nominalists argue that these objects are dependent on the practice of mathematicians, who make them up by conceptualizing them and introducing suitable mathematical principles. As a result, the same ontological criterion (ontological independence) yields conflicting outcomes regarding the ontology of mathematics: a positive answer from platonists, a negative one from nominalists (Bueno 2013).

As an alternative to (N1), and in contrast to (P1), one can adopt a form of agnosticism: mathematical objects need not be taken to exist in order for one to make sense of various aspects of mathematical practice. As noted, the issue of their existence, in the metaphysical sense of ontological independence, is left entirely open by the practice. Of course, it is still possible that mathematical objects do exist as platonists conceive of them: the view is not obviously incoherent. Given that the independent existence of abstract objects is not required by the practice nor is it ruled out, an agnostic attitude toward this issue is, thus, recommended. That is the attitude I adopt.

With regard to (N2), that is, if mathematical objects existed, they would be abstract, both nominalists and agnostics about mathematics can agree on. Mathematical objects, if they existed, would not be spatiotemporally located nor would they be causally active. They are not the kinds of things that are subject to these constraints. This point goes through whether one denies the existence of mathematical objects, as nominalists do, or whether one leaves this issue open, along an agnostic line.

With regard to (N3), mathematical objects are introduced via suitable mathematical principles (such as comprehension principles), and in this way they are dependent on the way they are described and thought of. Again, whether one is a nominalist or an agnostic about mathematics, this point still goes through.

With regard to (N4), mathematical statements even if true need not be necessarily so. As will become clear shortly, this feature emerges directly from the pluralism that is inherent in mathematics (and logic). After all, the truth of mathematical statements ultimately depends on the way in which they are formulated: the underlying mathematical framework used and the logical apparatus invoked. Such statements are not truth absolutely, but are relative. As a result, mathematical results are not necessary: they hold (or not) dependent on the underlying logic or the relevant mathematical framework. I will return to this point below.

This contigentism (or anti-necessitarianism) about mathematics is not, however, a trait shared by all nominalist views. For instance, Hartry Field's nominalism is still committed to mathematical theories being necessary, in a specific sense, even if they are not true. After all, on his view, these theories are conservative: they are consistent with every internally consistent claim about the physical world (Field 1980/2016, 1989). Conservativeness is a form of necessity given that it imposes consistency with *all* consistent claims regarding the non-mathematical world. First, it is a form of necessity as *invariability*, given that a conservative mathematical theory "describes matters that don't vary across [. . .] possible alternatives" (Kment 2014, p. 20), that is, however the physical world turns out to be, a conservative mathematical theory is consistent with it. Second, this is also a form of *unconditional* necessity, given that conservative mathematical theories "hold unconditionally—i.e., independently of what else is the case" (Kment 2014, p. 20), for no matter how the physical world ends up being, such theories are consistent with it too.

However, it is unclear that conservativeness is, in general, a good trait of a mathematical theory. After all, it may be useful to have mathematical theories that impose constraints (such as cardinality ones) on descriptions of the physical world, just as it is useful to have interpretations of physical theories that impose constraints (for instance, about what is going on beyond the appearances) on descriptions of the world. These constraints, in many instances, provide understanding about how the world could be in light of the relevant mathematical or empirical theories, even if they do not provide direct empirical information about the world.

Field's approach to nominalism shares the Quinean doctrine that indispensable quantification over mathematical objects requires their existence. This assumption is contested by both defensible forms of nominalism and agnosticism. This means that some device is needed to support the claim that quantification does not entail ontological commitment. On my view, the most straightforward way of achieving this is by adopting ontologically neutral quantifiers (Azzouni 2004, 2017; Bueno 2005).

It is important to note that quantification does not require the existence of the objects that are quantified over, even those that are indispensable to our best theories of the world. After all, in a statement such as, "Some sets are too big to exist", if the existential quantifier "some" is understood in an ontologically committing way, we end up with a contradiction: "There exist sets that do not exist". But this just fails to express the intended and perfectly correct claim to the effect that, among the sets, some do not exist (due to their size).

However, no such tension is found with ontologically neutral quantifiers. The central idea is that quantification and existence should be distinguished. Quantification indicates only whether the entire domain is within the scope of the quantifier or just a part of it is. Universal quantification concerns all of the domain (independently of the existence of the objects that are quantified over) and existential quantification concerns a portion of the domain (again, with no assumption regarding the existence of the relevant objects). In order to mark ontological commitment, an existence predicate is then introduced. In the case of the statement about sets above, it can be easily formulated with ontologically neutral quantification, as follows:

$$\exists x \ \ (Sx \wedge \neg Ex),$$

in which 'E' is an existence predicate and 'S' stands for *sets*.

In this way, one can quantify over mathematical objects, such as vectors in a Hilbert space, while formulating quantum mechanics, without thereby being committed to the existence of these objects. This is a stance that physicists recognize when it is insisted that Hilbert spaces are just part of the mathematics and not of the physics. Presumably, the ontological commitment in question is to the relevant physical traits rather than to what is described by the mathematical apparatus, despite the fact that such an apparatus is indispensable for the formulation of the physical theories under consideration. Given ontologically neutral quantifiers, quantification over indispensable objects, such as vectors in a Hilbert spaces, does not incur ontological commitment.

5.4 Mathematical Pluralism and Contingent Mathematics

Central to the case for the contingency of mathematical ontology is the plurality of mathematics. Classical mathematics establishes a number of dependences between, on the one hand, mathematical principles and, on the other, the framework in which these principles are formulated, the resulting theories, and the underlying logic that is adopted. The complex network of dependences that is generated questions any claim to the effect that mathematical results hold necessarily, given that the truth of mathematical statements varies with changes in any of these components.

5.4.1 Mathematical Pluralism

There are at least two kinds of pluralism in mathematics: pluralism about the framework used in the characterization of mathematical theories and pluralism about the formulation of mathematical theories themselves. A third kind of pluralism, about the underlying logic, is not restricted to mathematics per se, since changes in logics affect any theories that are formulated using them, whether the theories are mathematical or not. But given the significance of this pluralism to mathematics, it is important to consider it as well. Each will be examined in turn.

(a) *Mathematical framework pluralism*. Consider the different mathematical frameworks that can be used to formulate scientific theories. These different frameworks have very different expressive and inferential features, and dramatically change what can (or cannot) be derived from the theories in question.

For instance, quantum mechanics (whether relativist or not) can be formulated in frameworks as diverse as set theory, category theory, modal second-order logic, or second-order mereology. These frameworks provide very different resources to formulate the basic concepts and principles of the theory, construct models, and derive results from them. Given the difference in these resources, the resulting formulations need not be equivalent. In some cases, the resources will not be either logically or ontologically the same, despite being based on the same principles.

As an illustration, consider a formulation of quantum mechanics that, instead of adopting the more common Zermelo-Fraenkel set theory with the axiom of choice (ZFC), uses Quine's New Foundations set theory (NF) as its basis (Quine 1937). Formulations of quantum mechanics in terms of Hilbert spaces typically rely on the result that every Hilbert space has a base. But to obtain this result, the axiom of choice is required. It turns out, however, that the axiom of choice is incompatible with NF (Specker 1953). Thus, it cannot be assumed that every Hilbert space has a base and the resulting formulation of quantum mechanics in NF is substantially different from the one in ZFC, even though the same quantum mechanical principles are invoked in each case (see also Krause and Bueno 2007; da Costa et al. 2010). In this way, the underlying mathematical framework matters to the formulation of a theory, and various kinds of differences can emerge as distinct frameworks are employed.

(b) *Pluralism about the formulation of mathematical theories*. Another kind of pluralism is advanced in light of different formulations of the same mathematical theory. The differences, in this case, are not connected to the underlying framework used to formulate the theories in question: the theories themselves are formulated in distinct ways.

Consider, for instance, the different formulations of the calculus: the early formulation, articulated by Leibniz, invoked infinitesimals, but these entities were eventually abandoned, in the nineteenth century, when Cauchy and Weierstrass articulated a continuum that did not include these objects. Later in the twentieth century, infinitesimals were reintroduced, via model-theoretic resources, with Abraham Robinson's development of non-standard analysis (Robinson 1974). These theories have different ontologies, some include infinitesimals, others do not, and those that include such entities articulate distinct conceptions of these magnitudes. Despite all of these differences, basically the same theory of differential and integral calculus is advanced, at least in the sense that the same rules of calculation are advanced in each case.

As an additional example, consider, more generally, the different formulations of mathematical analysis: there is classical analysis (Rudin 1976), predicative analysis (Weyl 1918/1987), and constructive analysis (Bishop and Bridges 1985). These are inequivalent formulations of a major domain of mathematics. Different results are obtained, or not obtained in the case of certain classical theorems, and the results that are obtained are obtained differently, depending on the (predicativist or constructivist) constraints that are in place. Although the statement of the theorems may be verbally the same in certain cases, the mathematics is different, given the distinct methods of proof generation, such as the use of constructive proof procedures in establishing classical results (that is, results that have been originally obtained via non-constructive means). This marks a significant kind of plurality within mathematics, relative to the relevant concepts, methods, and proof procedures.

Consider also the different formulations of arithmetic, which can be characterized in Fregean terms (Hale and Wright 2001), or in terms of Peano axioms, set-theoretic models, or modal-structural interpretations (Hellman 1989), not to mention inconsistent arithmetic (Priest 1997; Priest 2000). Consider different formulations of set theory as well: ZFC, NF, NBG (von Neumann-Bernays-Gödel) are just three distinct formulations of set theory (see Fraenkel et al. 1973), some are finitely axiomatizable (such as NBG), others are not (such as ZFC); some are consistent with the axiom of choice (such as ZFC, assumed to be consistent), others are not (such as NF).

In all of these cases, different mathematical frameworks characterize differently the content of certain scientific theories, and mathematical theories do not have their own content uniquely determined, given the variety of different characterizations of these theories. Once again, the resulting theories are, in many instances, neither ontologically nor logically equivalent. The result is an additional source of pluralism within mathematics (see also Friend 2014 for further defense of mathematical pluralism).

(c) *Pluralism about logic*: Logical pluralism (see Beall and Restall 2006 as well as Bueno and Shalkowski 2009) provides an additional source of pluralism in mathematics, given that, by changing the underlying logic of a mathematical theory, different theorems follow (or fail to follow) from the same mathematical principles. The result is a different mathematical theory, with non-equivalent theorems and results, despite the fact that the same mathematical principles are invoked.

As an illustration, consider the fact that one cannot study the Russell set (the set of all sets that are not members of themselves) in classical set theories, whose underlying logic is classical, since such an inconsistent object trivializes the theories in question (that is, everything follows from them in that case). However, if the underlying logic is paraconsistent, it is perfectly possible to study the Russell set in a paraconsistent set theory: despite its inconsistency, it is still a perfectly coherent, non-trivial mathematical object (da Costa et al. 2007). As opposed to what happens in classical logic, in a paraconsistent logic the so-called principle of *Explosion* does not hold, that is, not everything follows from a contradiction (i.e., a statement of the form *A* and not-*A*). Thus, results about the Russell set that cannot be even stated, let alone proved, in classical set theories can be examined and established in a suitable paraconsistent set theory.

Something similar happens in constructive mathematics. As noted, certain results that can be proved in classical analysis by invoking non-constructive methods cannot be proved in constructive mathematics if more stringent constructive constraints are introduced. The same mathematical principles lead, in light of different logics, to mathematical theories that are significantly different. However, if constructive constraints are relaxed, for instance by allowing the same standards of definition of mathematical terms as those adopted in classical mathematics, but still changing the underlying logic to a constructivist one, more results from classical analysis can thereby be obtained (Bishop and Bridges 1985). Despite that, different underlying logics still end up yielding different mathematics.

Given the dependence of mathematical results on the mode of introduction of mathematical objects (or on the definition of mathematical terms) and the underlying logic that is adopted, it is unclear that these results are ultimately necessary. They depend on the particular logic and the specific modes of definition that are employed. Mathematical results, thus, vary across possible alternatives: different logics, different modes of definition, different ways of introducing mathematical terminology all affect such results. Their truth is not unconditional: it depends on each of these salient traits of mathematical practice. Nor are mathematical theorems true in every possible world, given that they *fail* in some contexts, such as, as will become clear below, the lack of well-ordering of sets in the absence of the axiom of choice.

5.4.2 Contingent Mathematics

In light of mathematical pluralism, it is unclear that mathematical statements if true are necessarily true. The truth of any such statements is relative to a particular mathematical context (the relevant mathematical principles and the underlying logic). If such principles change, the truth value of the relevant statements may change as well. Consider, for instance, the question: Is every set well ordered? The answer depends on whether the axiom of choice holds or not. If it does, then the answer is affirmative; otherwise, it is not. The necessity that is obtained in mathematics is, at best, a *conditional necessity*: assuming certain mathematical principles (particular mathematical concepts and their relations) and assuming a given logic, the result in question holds. If one adopts both the axiom of choice and classical logic, it then follows that every set can be well-ordered; otherwise, this is not the case.

Note that it is not up to us which results hold or fail to hold, assuming particular principles and a given logic. This supports the objectivity of mathematics, despite mathematical and logical pluralism. This objectivity is often confused with the perceived unconditional necessity of mathematics, which yields the illusion that mathematical statements if true are necessarily so. The apparent necessity surfaces by ignoring the variety of mathematical frameworks, the multitude of formulations and characterizations of mathematical theories, and the multiplicity of underlying logics. Once these pluralisms are acknowledged, the lack of necessity of mathematical theories, and the corresponding contingency of mathematical objects, become manifest. Mathematical results do not hold in general, but are dependent on these traits, which, in turn, need not hold in general either. In the end, mathematical statements are not necessary, but only hold contingently. I will return to this point below.

Furthermore, in light of mathematical and logical pluralism, none of the conceptions of necessity examined above applies to mathematics either. I will consider each in turn.

(a) *Necessity as invariability*. According to this conception, "[a] truth is necessary if it describes matters that don't vary across [. . .] possible alternatives" (Kment 2014, p. 20). However, mathematical statements *do* vary across "possible alternatives". A mathematical statement that is true in a given (mathematical) framework is false in another, such as the case of the well-ordering principle, which holds only if the axiom of choice does. The truth of a mathematical statement also depends on the theory in which it is formulated: the Russell set does not exist in classical set theories (assumed to be consistent), but it does in a paraconsistent set theory (assumed to be non-trivial). Mathematical results often depend on the logic that is used to formulate them: the predicativist continuum is importantly different from the classical one, just as the continuum with infinitesimals differs from those that do not include them.

(b) *Necessity as unconditional truth*. Recall the conception of necessity according to which, "[t]he metaphysical necessities are those propositions that hold unconditionally—i.e., independently of what else is the case. By contrast, a true proposition is a *contingent* truth if its truth is contingent on something else, i.e. dependent on what else is the case" (Kment 2014, p. 20). Mathematical statements are then contingent: their truth is contingent on the underlying mathematical framework in which they are formulated, on the particular formulation of the results in question, and on the relevant logic that is adopted. Every model of analysis is isomorphic provided that second-order logic is invoked; otherwise, in a first-order context, there are non-standard models of analysis. The categoricity of real analysis depends on the underlying logic; it is not, thereby, necessary.

(c) *Necessity as truth in all possible worlds* (Lewis 1986). As we saw, on a possible world conception of necessity, a mathematical statement is necessary provided that it is true in *all* possible worlds. Given that mathematical objects are abstract and Lewisian worlds are concrete, there is nothing in these worlds that could undermine the truth of a mathematical statement. Hence, or so the argument goes, all true mathematical statements are necessarily true.

But this is not quite right. The account fails to accommodate the dependence of the truth of a mathematical statement on the framework in which the statement is formulated. For instance, assuming the axiom of choice, first-order logic is complete; otherwise, it is not. The completeness of first-order logic (a model-theoretic, set-theoretic result) is dependent on the axiom of choice (in the form of Zorn's lemma). Hence, the result does not hold in general: the completeness fails in every world in which the axiom of choice also fails (for instance, in all worlds in which constructive features are in place). Thus, the result is not necessary.

What is necessary, one could argue, is something else: If the axiom of choice is assumed, together with classical set theory and classical logic, then first-order logic is complete. But the necessity of the completeness of first-order logic does not follow unless one assumes the necessity of the axiom of choice and of classical logic. However, none of them is necessary either: the axiom of choice fails in constructive contexts and classical logic fails in inconsistent, incomplete domains (Bueno and Shalkowski 2009). In the end, what results is a conception according to which mathematics, including its objects, is contingent.

It may be objected that a given mathematical statement may be true relative to a particular framework, theory or logic, and false relative to others. However, this does not provide reasons to believe that the proposition expressed by the statement in question, relative to the framework, theory or logic in question, is contingent. After all, in each case, different propositions are being expressed, for they are propositions about distinct mathematical structures, formulated in distinct theories, or in different logics. (I owe this objection to an anonymous reviewer.)

In response, the objection grants the crucial point to contingentism: the relevant mathematical statements, being relative to particular frameworks, theories or logics, are not true in general, given that they fail to hold in different frameworks, theories or logics. In the end, there does not seem to be absolute necessities in mathematics, only relative ones: necessities given a framework, a theory, or a logic. But given that the frameworks, theories and logics in question, in turn, are not necessary either (for the reasons discussed earlier in this work), the conditional necessities invoked do not go very far. After all, one cannot assert the necessity of the frameworks, theories and logics in question to obtain the necessity of the relevant mathematical statements. As a result, the relative necessities cannot be discharged, and contigentism seems to go all the way down.

Finally, note that the objectivity of mathematics can be accommodated here without any commitment to the existence of mathematical objects: given certain mathematical principles and a logic, what follows from such principles does not depend on us; it is simply a matter of the logical relations between the principles in question and what logically follows from them. Thus, mathematical objectivity, as opposed to what happens in platonism, does not depend on the existence of mathematical objects.

5.5 Conclusion

Platonism provides an interpretation of mathematics that adds to mathematical practice two substantial metaphysical assumptions: (i) mathematical objects exist (in the sense that they are mind and language independent entities), and (ii) mathematical statements if true are necessarily true (in the sense that they are invariably and unconditionally true or true in all possible worlds). These assumptions are not needed to make sense of mathematical practice, and they make it much harder to accommodate significant forms of pluralism that are found in that practice: pluralism about the formulation of mathematical theories, about their interpretation, and about the underlying logics and frameworks invoked in mathematics.

Once such pluralisms are recognized, it becomes clear that the (apparent) unconditional necessity of mathematics is ultimately illusory, and with the use of neutral quantifiers, no commitment to the existence of mathematical objects is forthcoming. The result is a less metaphysically loaded account of mathematics that openly embraces the contingency of mathematical objects, while preserving the objectivity of mathematics.

Acknowledgements My thanks go to Pablo Alcuña Luongo, José Tomas Alvarado, Sam Cowling, Diego Romero, Cristian Soto, and audiences at Denison University and Pontificia Universidad Católica de Chile for helpful discussions of earlier versions of this work. Thanks also go to an anonymous reviewer for very helpful comments that led to significant revisions in the paper. Many thanks are also due to Concha Martínez Vidal and José Luis Falguera Lopez for all their help, patience, and support during the writing of this article.

References

Azzouni, J. (1994). *Metaphysical myths, mathematical practice: The ontology and epistemology of the exact sciences*. Cambridge: Cambridge University Press.

Azzouni, J. (2004). *Deflating existential consequence*. New York: Oxford University Press.

Azzouni, J. (2017). *Ontology without borders*. New York: Oxford University Press.

Balaguer, M. (1998). *Platonism and anti-platonism in mathematics*. New York: Oxford University Press.

Beall, J. C. (1999). From full-blooded platonism to really full-blooded platonism. *Philosophia Mathematica, 7*, 322–325.

Beall, J. C., & Restall, G. (2006). *Logical pluralism*. Oxford: Oxford University Press.

Bishop, E., & Bridges, D. (1985). *Constructive analysis*. Berlin: Springer.

Bueno, O. (2005). Dirac and the dispensability of mathematics. *Studies in History and Philosophy of Modern Physics, 36*, 465–490.

Bueno, O. (2011). Logical and mathematical knowledge. In S. Bernecker & D. Pritchard (Eds.), *Routledge companion to epistemology* (pp. 358–368). London: Routledge.

Bueno, O. (2013). Nominalism in the philosophy of mathematics. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*. (Fall 2013 edition). http://plato.stanford.edu/archives/fall2013/entries/nominalism-mathematics/

Bueno, O., & Shalkowski, S. (2009). Modalism and logical pluralism. *Mind, 118*, 295–321.

Bueno, O., & Shalkowski, S. (2015). Modalism and theoretical virtues: Toward an epistemology of modality. *Philosophical Studies, 172*, 671–689.

da Costa, N. C. A., Krause, D., & Bueno, O. (2007). Paraconsistent logics and paraconsistency. In D. Jacquette (Ed.), *Philosophy of logic* (pp. 791–911). Amsterdam: North-Holland.

da Costa, N. C. A., Krause, D., & Bueno, O. (2010). Issues in the foundations of science, I: Languages, structures, and models. *Manuscrito, 33*, 123–141.

Field, H. (1980/2016). *Science without numbers: A defense of nominalism.* Princeton: Princeton University Press (Second, expanded edition, Oxford: Oxford University Press, 2016).

Field, H. (1989). *Realism, mathematics and modality*. Oxford: Basil Blackwell.

Fraenkel, A., Bar-Hillel, Y., & Levy, A. (1973). *Foundations of set theory*. (Second revised edition with the collaboration of D. van Dalen.) Amsterdam: Elsevier.

Friend, M. (2014). *Pluralism in mathematics: A new position in philosophy of mathematics*. Dordrecht: Springer.

Gödel, K. (1964). What is cantor's continuum problem? In P. Benacerraf & H. Putnam (Eds.), *Philosophy of mathematics: Selected readings* (2nd ed., pp. 470–485). Cambridge: Cambridge University Press.

Hale, B., & Wright, C. (2001). *The reason's proper study: Essays towards a Neo-Fregean philosophy of mathematics*. Oxford: Clarendon Press.

Hellman, G. (1989). *Mathematics without numbers: Toward a modal-structural interpretation*. Oxford: Clarendon Press.

Kment, B. (2014). *Modality and explanatory reasoning*. Oxford: Oxford University Press.

Krause, D., & Bueno, O. (2007). Scientific theories, models, and the semantic approach. *Principia, 11*, 187–201.

Lakatos, I. (1976). *Proofs and refutations: The logic of mathematical discovery* (J. Worrall & E. Zahar, Eds.). Cambridge: Cambridge University Press.

Lewis, D. (1986). *On the plurality of worlds*. Oxford: Blackwell.

Priest, G. (1997). Inconsistent models of arithmetic, part I: Finite models. *Journal of Philosophical Logic, 26*, 223–235.

Priest, G. (2000). Inconsistent models of arithmetic, part II: The general case. *Journal of Symbolic Logic, 65*, 1519–1529.

Quine, W. V. O. (1937). New foundations for mathematical logic. *American Mathematical Monthly, 44*, 70–80.

Resnik, M. (1997). *Mathematics as a science of patterns*. Oxford: Oxford University Press.

Robinson, A. (1974). *Non-standard analysis*. Amsterdam: North Holland.
Rudin, W. (1976). *Principles of mathematical analysis* (3rd ed.). New York: McGraw-Hill.
Shalkowski, S. (1994). The ontological ground of the alethic modality. *The Philosophical Review, 103*, 669–688.
Shapiro, S. (1997). *Philosophy of mathematics: Structure and ontology*. New York: Oxford University Press.
Specker, E. (1953). The axiom of choice in Quine's new foundations for mathematical logic. *Proceedings of the National Academy of Science, 39*, 972–975.
Weyl, H. (1918/1987). *The continuum: A critical examination of the foundations of analysis* (Translated from the German by Stephen Pollard and Thomas Bole. The first German edition was published in 1918). New York: Dover.

Chapter 6
Is There a Fact of the Matter About the Existence of Abstract Objects?

Mary Leng

Abstract In 'Empiricism, Semantics, and Ontology', Rudolf Carnap argues that the ontological questions that philosophers attempt to answer are ill-formed. Internally to the frameworks of conventions that give meaning to our terms, questions like 'Are there numbers?' have straightforward and trivial answers: the axioms for number theory tell us how we are to use number terms, and the existence of natural numbers follows trivially from these axioms. On the other hand, if we try to step outside of these meaning-giving frameworks and ask 'Are there *really* numbers?', having set aside the meaning-giving conventions that characterize the natural number framework, this external question has been given no sense. All we can ask is the pragmatic question of whether the natural number framework is a useful one for us to adopt. Quine's naturalistic approach to ontology offers a way out of Carnapian ontological skepticism, which proceeds by questioning Carnap's strict practical/evidential divide, arguing that the practical decision to adopt a framework which speaks of Fs in our best scientific theories just is evidence for the existence of Fs. This chapter considers some neo-Carnapian challenges to Quine's naturalism and argues that a recent challenge from Penelope Maddy (Defending the axioms: on the philosophical foundations of set theory. OUP, Oxford, 2011) is right in suggesting that at least some ontological questions do not have deep answers.

Keywords Carnap · Quine · Azzouni · Yablo · Maddy · Naturalism · Realism · Ontology

6.1 Introduction

As Quine (1948, p. 1) famously noted, the "curious thing about the ontological problem is its simplicity. It can be put in three Anglo-Saxon monosyllables: 'What is there?' It can be answered, moreover, in a word—'Everything'—and everyone

---

M. Leng (✉)
Department of Philosophy, University of York, York, UK
e-mail: mary.leng@york.ac.uk

J. L. Falguera, C. Martínez-Vidal (eds.), *Abstract Objects*, Synthese Library 422,
https://doi.org/10.1007/978-3-030-38242-1_6

will accept this answer as true." Despite its apparent simplicity, however, there remains room for (what Quine again called) "disagreement over cases". What gets to be included under 'everything'? The medium-sized physical objects we take ourselves to observe? The unobservable physical objects posited by our scientific theories? Non-physical 'spiritual' entities such as Gods or souls? Abstract objects such as numbers or propositions? Does every*thing* include *universal* as well as *particular* things? Moral objects such as normative reasons to act? Unactualized possibilities? Fictional characters? The 'cases' are where all the action happens, and the difficulty in deciding the issue for these various cases is seemingly what has kept this philosophical debate rumbling on.

There is, however, a more skeptical diagnosis of why this seemingly simple question has proved so recalcitrant. According to the skeptical diagnosis, the debate rumbles on not because it is *difficult* to settle the cases and agree what falls under 'everything', but because it is *impossible* to do so: the philosophers' ontological question, despite its apparent simplicity, has not been given a sense. This was the famous complaint of Carnap (1950), who thought that questions of what there is could only be answered relative to an assumed framework of stipulated meaning-giving linguistic conventions, and that such 'internal' existence questions could often be answered trivially. Where metaphysicians try to answer the framework-independent *external* questions, the only meaningful question they can ask is the practical one of whether a given framework is useful to adopt. Quine's approach to ontology was presented as a non-skeptical response to Carnap's ontological skepticism; while Quine agreed with much of Carnap's critique of traditional metaphysics, he argued that the practical decision to adopt one framework over another in our best attempts to organize our experience should be taken as evidence that the conceptual scheme offered by the framework has things right about the world we experience. As such, Quine argued, we ought to accept the existence of objects that exist according to the internal standards of those frameworks that have proved to be practically most useful in describing, explaining, and predicting our experiences. The upshot of this is Quine's naturalistic approach to ontology, according to which we are simply to read our ontological commitments off our current best scientific theory.

Despite Quine's rejoinder to Carnap, and despite the prominence of the Quinean approach to ontology ever since, skeptical concerns about the ontological project are again on the rise. In recent years there has been a resurgence of work on 'metametaphysics', with the question of whether ontological questions can be answered taking central ground (see, e.g., the papers in Chalmers, Manley, and Wasserman 2009). I would like to discuss two relatively early precursors to this early twenty-first century revival of ontological skepticism, offered by Jody Azzouni (1998) and Stephen Yablo (1998). I will argue that, to the extent that Yablo and Azzouni offer challenges to the Quinean naturalist approach to ontology, these challenges can be responded to with a more thoroughgoing naturalism. It is this approach that stands behind my (2010) defense of mathematical fictionalism. However, more recently Penelope Maddy (2011) has offered a further route to ontological skepticism that may threaten even this more thoroughgoing naturalistic project. I will consider Maddy's concerns, and will concede to Maddy that the Quinean naturalist project may indeed leave the answer to some questions concerning the existence of abstract objects "a matter of decision rather than assertion" (Carnap 1950).

6.2 Abstract Objects: The Quinean Way

Our concern in particular is with abstract objects, but our discussion must start with the ontological problem generally, and in particular in the question of how to go about dealing with the 'cases' over which there is disagreement – that is, working out what falls under 'everything'. The background to the Quinean approach to ontology starts, as I have said, with Carnap, and in particular with Carnap's classic paper, 'Empiricism, Semantics, and Ontology' (Carnap 1950). Carnap presents his discussion of ontology in this paper as an aid to empiricists who wish to "overcome nominalistic scruples" (Carnap 1950, p. 206) that lead them to shy away wherever possible from talk of abstract objects. Carnap argues, in ESO, that using linguistic forms in one's theorizing that purport to refer to objects of a particular sort does not carry with it any metaphysical commitment to the existence of the objects to which our terms appear to refer. However, while one might think that this is good news for nominalists who may wish, e.g., to use terms that purport to refer to abstract mathematical objects in their theorizing without incurring a commitment to the existence of abstract objects, Carnap sees his argument not as vindicating nominalism, but as a step to showing that no sense can be given to the debate between realism and nominalism, since no sense can be made of ontological questions of a metaphysical bent.

Thus, in considering how to deal with ontological questions, Carnap introduces the notion of a linguistic framework, and notes that questions of the form "Do *F*s exist?" may be asked from within a particular linguistic framework, or alternatively from some perspective external to that framework. As an internal question, "Do numbers exist?", for example, just asks whether the rules governing assertion within the relevant linguistic framework (in this case, the framework of arithmetic) allow us to assert the existence of numbers. Given that the framework of linguistic rules that give meaning to our number talk will include the Dedekind-Peano axioms, and given that these axioms trivially imply the existence of numbers, as an internal question, "Do numbers exist?" has a trivial and positive answer. However, according to Carnap, one's acceptance of this answer "does not imply embracing a Platonic ontology" (Carnap 1950, p. 206). Acceptance of a linguistic framework implies no more than acceptance of a way of talking, with no corresponding belief implied about the independent existence of those things to which our language appears to allow us to refer. Similarly, within what Carnap calls the "thing" framework (roughly, the framework of physical objects), we may ask "Do tables exist?", and answer this internal question in the affirmative (as a result of a combination of meaning-giving linguistic conventions, and empirical observation), but again this affirmative answer does not imply any substantial belief about the independent existence of tables.

If we wish to ask "Do numbers *really* exist?" (or indeed "Do tables *really* exist?"), meaning by this something more than "Do numbers exist according to the framework of Peano arithmetic" (or "Do tables exist according to the framework of things"), then, according to Carnap we are attempting to ask an external question, about the linguistic framework itself. And, once we have set aside the conventional framework rules that give meaning to existence claims within a framework, the only meaningful question left for us to ask is, in Carnap's view, a practical one: is this linguistic framework appropriate to the task for which we would like to use it? Again, a positive answer to this does not imply Platonism; merely saying that a way of speaking is suited to some particular task in no way commits us to a further metaphysical claim about the objects about which it allows us to speak.

It is important to be clear, however, that Carnap's repudiation of Platonism does not amount to a commitment to nominalism. For, while Carnap and nominalists alike claim that our use of, e.g., mathematical language need not commit us to realism with respect to mathematical objects, Carnap goes further to add that there are *no* cases where one can move from the use of a language to the reality of its objects. Carnap clarifies this point in his 'Intellectual Biography' as follows:

> ...if one proceeds from the discussion of language forms to that of the corresponding metaphysical theses about the reality or unreality of some kind of entities, he steps beyond the bounds of science. Carnap (1963), p. 19

Science does not give a meaning to metaphysical doctrines about the reality or unreality of the objects to which it refers, and so, for Carnap, these doctrines are pseudotheses (Carnap 1963, p. 50). Thus, Carnap tells us, "If someone decides to accept the thing language, there is no objection against saying that he has accepted the world of things. But this must not be interpreted as if it meant his acceptance of a *belief* in the reality of the thing world; there is no such belief or assertion or assumption, because it is not a theoretical question." (Carnap 1950, pp. 207–8) We may talk about 'things', form beliefs about 'things', assert truths about 'things', but, for Carnap, none of this commits us to the thesis of the existence of *any*thing.

Quine's response to Carnap's ontological scepticism is well known. While Quine agrees with Carnap and the positivists that the 'metaphysical' doctrines of realism and nominalism, *when developed from the perspective of 'first philosophy'*, are pseudo-theses, this does not, he thinks, imply that there is *no* interesting interpretation of these ontological doctrines. Quine's own presentation of their difference suggests that the issue is merely linguistic:

> Now if [Carnap] had a better use for this fine old word 'ontology', I should be inclined to cast about for another word for my own meaning. But the fact is, I believe, that he disapproves of my giving meaning to a word which belongs to traditional metaphysics and should therefore be meaningless. Now my ethics of terminology demand, on occasion, the avoidance of a word for a given purpose when the word has been preempted in a prior meaning; meaningless words, however, are precisely the words which I feel freest to specify meanings for. Quine (1951), p. 126.

However, Quine goes on to indicate that there is a more substantial difference at issue here, suggesting that his

> adoption of the word 'ontology' for the purpose described is not as arbitrary as I make it sound. Though no champion of traditional metaphysics, I suspect that the sense in which I use this crusty old word has been nuclear to its usage all along. (Quine 1951, pp. 126–7)

Quine thus supposes, *contra* Carnap, that there is a meaningful core to the traditional ontological doctrines, which is brought out in his own method of recasting ontological questions.

And what is Quine's method for understanding such questions? Well, cast in Carnapian terminology, Quine considers the 'framework' of natural science, and argues that the 'external' ontological questions we wish to ask with respect to this framework are answered precisely by the answers to the framework's 'internal' ontological questions. If we want to ask whether *F*s *really* exist, we simply ask whether, according to the framework adopted for the purpose organising and predicting our experience – i.e. the framework of natural science – *F*s exist. That is, we ask whether, according to the framework of natural science, we are justified (by the framework's internal standards) in making assertions that existentially quantify over *F*s. An answer in the affirmative commits us to realism with respect to *F*s.

Quine's defence of this position consists of an argument to the effect that Carnap's internal/external distinction is untenable. In particular, Quine attacks the internal/external distinction as just being another version of the ill-fated analytic/synthetic distinction. It is central to Carnap's claim that many 'internal' existence questions are trivially answered given the framework rules, and thus are not properly metaphysical, that the framework rules have the status of *conventions*. For Carnap, we start out with meaning-giving conventions and can *only then* begin empirical inquiry, so that our empirical questions only have answers relative to a collection of untestable conventional assumptions. Quine's attack on the analytic/synthetic distinction is a direct attack on this divide: in the light of recalcitrant evidence we can just as much question the framework-rules (pick a new framework) as question empirical claims made against the backdrop of those rules. Thus, for Quine, the *practical* decisions we make to adopt one framework over another are just as responsive to empirical evidence as the theoretical observations we make in light of experience relative to the framework rules. Practical reasons to adopt a linguistic framework in describing, predicting, and explaining our experiences just are reason to believe the claims of the framework, 'conventional' rules and all.

The upshot of all this is that Quine is able to argue that when we ask 'metaphysical' questions regarding what there is, we are effectively asking, 'What sorts of objects does our framework of natural science require us to say that there is?', where a framework says that there are *F*s if it licenses existential quantifications over *F*s (that is, if it supports inferences to claims of the form 'there are *Fs*'). This ontological approach consists, as Jody Azzouni (1998) observes, of two components, "a "criterion for what exists" (CWE)", where according to Quine, this is given by looking to the commitments of our best scientific theories, "and a "criterion for recognizing what a discourse commits us to" (CRD)" (Azzouni

1998, p. 2), where, for Quine, this is a matter of uncovering the quantifier-commitments of that discourse (i.e., the existentially quantified claims licensed by the discourse). Given Quine's CWE and his CRD, the indispensability of quantification over mathematical objects in our best scientific theories means that, despite what 'nominalistic scruples' we may have begun with, if we wish to use natural science as it stands, then we are committed to the existence of the abstract mathematical objects quantified over those theories.

6.3 Naturalistic Scruples About Quine's Criterion of Ontological Commitment

The claim that there is something wrong with taking quantifier commitment to signify ontological commitment could be argued for on many grounds, but given the naturalistic setting of Quine's approach to ontology, perhaps the most convincing attack has come from within the naturalistic programme. Penelope Maddy (1992) opened the doors for a naturalistic rejection of Quine's account, arguing that if we agree with the Quinean naturalist to take our philosophical cues from scientific practice (rather than imposing the conclusions of *a priori* philosophical reasoning on scientists), then it is difficult to defend the claim that ontological commitment corresponds to quantifier commitment in light of scientists' attitudes to their quantifications. Taking a closer look at scientific practices, Maddy points out that many examples can be found which suggest that scientists do not always see the indispensable occurrence of quantification over *F*s in their theories as requiring commitment to the existence of *F*s. With respect to theories as a whole, Maddy points out that the indispensable use of some theories does not stand in the way of scientists' taking an instrumental attitude to those theories; while with respect to individual objects referred to in a theory, Maddy's suggestion is that the indispensable occurrence of an object in a theory is not by itself enough to convince scientists of the existence of that object. Thus, it seems, Quine's criterion for recognising the ontological commitments of a discourse (CRD) falls foul of his naturalistic (CWE) – if we are going to allow science to tell us what there is, then Maddy's conclusion is that we had better listen a lot more carefully than Quine does to discover what it is that science is telling us.

Maddy's evidence for the first claim – that some theories are indispensable to our best science and yet considered instrumentally, rather than as literally true – comes from noticing that there are some theories that are indispensable to our best science and yet, uncontroversially, are literally false. Science is full of literally false, but instrumentally useful, theoretical assumptions. The mere presence of the assumption of an infinitely deep ocean in our best theory of water waves, for example, should not convince us that such an object really exists (Maddy 1997, p. 143). Now some of these falsehoods are easily dealt with by Quine, who points out that they can be reinterpreted as shorthand for literally true theories about the behaviour of phenomena as they approach limits. Our best science contains these literally true theories, for which the idealized falsehoods are a useful shorthand. But not all of these false theories are dealt with so easily. Thus, for example, in the case of fluid dynamics it's unclear how the assumption that fluids are continuous could be replaced with claims about the behaviour of real fluids as they approach a continuous limit. After all, Maddy points out, "fluid dynamics isn't more applicable to one fluid than another, depending on how closely that fluid approximates a continuum; rather, it provides a workable account of any fluid" (Maddy 1997, p. 145). It is difficult to see how such an idealization could be reinterpreted as a literally true theory. But if no such reinterpretation can be given, it seems that a distinction must be made in natural science between those of our theories that are taken to be literally true, and those that are useful instruments that are literally false. Thus we have a first reason, from the perspective of naturalism, to challenge Quine's CRD.

A second reason for challenging the identification of ontological commitment with quantifier commitment comes from noticing that scientists often remain sceptical about the existence of objects quantified over in their theories even when they have no reason to think of their theories as literally false. In support of this claim, Maddy points to the widespread scepticism regarding the actual existence of atoms, circa 1900, even though atomic theory was an impressively successful part of the science of the day. In this case, most scientists waited for some form of 'direct evidence' of the existence of atoms, which came in the form of Perrin's Brownian motion experiments. It seems, then, that scientists often reserve judgement regarding the existence of theoretical entities until they can have some direct evidence of their existence, such as the observation of events that are best explained as having been directly caused by the entities in question.

We thus have two aspects of scientific practice that suggest that scientists do not take their theories' quantifier commitments as always indicative of their own ontological commitments as users of those theories. Scientists may see a theory in its entirety as of merely instrumental use, rather than as literally true; and they may reserve judgements as to the existence of the entities referred to in their theories until they have more direct evidence of their existence, via the observation of events which would be difficult to explain without invoking the entities in question as causes. From these considerations, we may suggest two restrictions on our CRD: firstly, it should not commit us to the existence of those objects quantified over in parts of our scientific theory that are not understood as literally true, and secondly, it should not commit us to the existence of those objects quantified over in our theory for which we do not have any more direct causal evidence.[1]

These two restrictions look promising from the perspective of nominalists who wish to draw a line between the abstract mathematical objects and the theoretical, but physical, objects referred to in our theories. Using the 'literal theory' restriction,

[1]Note that each of these restrictions should be taken as capable of standing alone – arguments against the 'causal evidence' restriction, for example, need not impinge on our acceptance of the 'literal theory' restriction, and vice versa.

we may deny the existence of mathematical objects that are indispensably quantified over in our theories if we can show that they occur in those parts of our theories that are to be interpreted instrumentally, rather than literally. Alternatively, using the 'causal evidence' restriction, we may deny the existence of mathematical objects occurring in our theories if we have no direct evidence in the form of phenomena that can only be explained as having been caused by these objects. The latter restriction, in particular, would seem to be an easy way of ruling out commitment to any mathematical objects: since these objects are usually considered abstract and acausal, the sort of direct causal evidence the second restriction requires would never seem to be available for the objects referred to in our mathematical theories. Nevertheless, despite these nominalistic hopes, in two papers published in 1998 both Stephen Yablo and Jody Azzouni presented arguments to the effect that, once one realises that Quine's CRD can be challenged, the upshot is not a new hope for nominalism, but a return to Carnapian ontological scepticism.

6.4 Does Ontology Rest on a Mistake?

6.4.1 Yablo's Critique – Difficulties with the Metaphorical/Literal Distinction

Stephen Yablo's challenge to the Quinean ontological project begins with an argument that the Quinean naturalist must be able to separate the literal from the non-literal parts of her theory prior to drawing ontological conclusions. Yablo presents Quine as advising us to

> (Q) count a thing as existing iff it is a commitment of your best theory, i.e., the theory's truth requires it. (Yablo 1998, p. 245)

However, Yablo notes, this is lousy advice if it is the case that parts of our best theory are best thought of as non-literal:

> What though if my best theory contains elements *S* that are there not because they are such very good things to believe but for some other reason, like the advantages that accrue if I *pretend* that *S*? Am I still to make *S*'s commitments my own? One certainly hopes not; I can hardly be expected to take ontological guidance from a statement I don't accept, and may well regard as false!
> (Yablo 1998, p. 245)

If our best theory contains non-literal parts, then we must revise (Q) in such a way that it counsels a different approach to those non-literal aspects of our theory. But in order to do this, we first need to be able to identify those parts of our best theory that are to be taken literally, and separate them from the non-literal parts. It is the claimed impossibility of this task that leads Yablo to conclude that the Quinean project of ontology "rests on a mistake". In short, Yablo's argument

> . . . goes like this: To determine our commitments, we need to be able to ferret out all traces of non-literality in our assertions. If there is no feasible project of doing *that*, then there is no feasible project of Quinean ontology. (Yablo 1998, p. 233)

Much of the work of Yablo's paper therefore is taken up with defence of the claims first that our best theories are likely always to include non-literal aspects, and second, that no easy line can be drawn between the literal and the non-literal parts of our theories.

Yablo explicitly presents his objection to the Quinean ontological project as a revival of a form of Carnapian ontological scepticism. In particular, Yablo argues that the metaphorical/literal distinction can be seen as a version of something like the internal/external distinction that survives Quine's critique. Quine's strategy for rejecting the internal/external distinction is to argue that the cogency of such a distinction rests on another ill-fated dichotomy, the distinction between analytic and synthetic truths. While some of Quine's arguments seem off the mark here,[2] Yablo concedes that Quine's critique hits the mark as a critique of the claim that linguistic frameworks can be thought of as picked out by analytic rules that determine the meanings of statements made within the framework. But this is not, according to Yablo, enough to force the abandonment of *any* useful notion of a linguistic framework. Identifying three steps in Quine's critique of Carnap, Yablo presents an alternative notion of a linguistic framework that is not touched by these three steps:

> Look again at the three stages. The first tells us that frameworks are not to be seen as sole determinants of meaning. All right, let '*X*''s meaning depend on factors that the framework has no idea of; let '*X*' have its meaning quite *independently* of the framework. The second tells us that the rules about what to say when had better not be rules about what to believingly assert. All right, let them be rules about what to *put forward*, where this is a conversational move falling short of assertion. The third tells us that if frameworks are non-doctrinal, this is not because they are adopted for reasons like simplicity, fruitfulness, and familiarity. All right, let the conclusion be reached by another and more direct route; let us identify frameworks outright with practices of such and such a type, where it is independently obvious that to engage in these practices is not thereby to accept any particular doctrine.
>
> Now, what is our usual word for an enterprise where sentences are put at the service of something other than their usual truth-conditions, by people who may or may not believe them, in a disciplined by defeasible way? It seems to me that our usual word is 'make-believe game' or 'pretend game'. Make-believe games are the paradigm activities in which we 'assent' to sentences with little or no regard for their actual truth-values. (Yablo 1998, p. 243)

The notion of a linguistic framework is thus resurrected by Yablo in his discussion via the concept of a make-believe game.

The distinction between internal existence questions and external existence questions of a philosophical kind – i.e. between the internal existence questions, which Carnap takes to be trivial and non-philosophical, and the external existence

[2] See, e.g., Bird (1995) for a close look at some problems with Quine's critique.

questions that go beyond questions regarding which frameworks are practically useful, which Carnap takes to be meaningless – is reprised in Yablo's picture in the distinction between questions as to what is merely true-in-a-story and questions as to whether a particular story is really (literally) true. Quine's blurring of the distinction comes down to the claim that all of the linguistic frameworks used in science are to be taken as literally true, since for Quine practical reasons to use these frameworks *just are* reasons to believe their assumptions. However, given the apparent indispensability of non-literal forms of language in our current best science, this Quinean move seems unacceptable. If theoretical fictions such as idealizations are indispensable in science, then we cannot always take a practical reason to speak 'as if' there are *F*s in the context of our best scientific theories as a reason to believe that there *really* are *F*s.

Quine of course recognises the presence of non-literal components such as explicit idealizations in our ordinary scientific theories. Nevertheless, he assumes that in our *best* theories these idealizations will be eradicated, at least in the long term. In response, Yablo argues, this line of defence is not open to Quine in bolstering his CRD, since, Yablo claims, we have no way of recognizing when a theory has eradicated its non-literal parts prior to having an answer to the ontological issue that concerns us:

> [Quine's] advice is to countenance numbers iff the *literal* part of our theory quantifies over them; and to count the part of our theory that quantifies over numbers literal iff there turn out to really be numbers. (Yablo (1998), p. 258)

If Quine's CRD is must distinguish between the literal and the non-literal parts of a theory, then in the absence of a separate test for what is to count as literal, he seems stuck with a vicious circle.

Yablo sums up his critique of Quine as follows:

> Quine's idea was that our ordinary methods could be 'jumped up' into a test of literal truth by applying them in a sufficiently principled and long-term way. I take it as a given that this is the one idea with any hope of attaching believable truth values to philosophical existence-claims. Sad to say, the more controversial of these claims are equipoised between literal and metaphorical in a way that Quine's method is powerless to address. It is not out of any dislike for the method—on the contrary, it is because I revere it as ontology's last, best hope—that I conclude that the existence-questions of most interest to philosophers are moot. If they had answers, (Q) would turn them up. It doesn't, so they don't. (Yablo 1998, pp. 259–60)

The negative conclusion we are left with, then, is that in the absence of a clear way of drawing the metaphorical/literal distinction, there is no hope for the Quinean ontological project. And given that this is the only such project that has looked like it stands any chance at success, there seems to be no hope whatsoever for the philosophical project of ontology.

6.4.2 Azzouni's Critique – The Impossibility of Arguing for a CRD

Jody Azzouni's alternative critique of the Quinean ontological project comes from considering how one might go about trying to defend one's choice of CRD. Against Quine's CRD which identifies quantifier commitment with ontological commitment, Azzouni considers an alternative CRD in which ontological commitment is indicated by the use of an existence predicate. "For example," Azzouni tells us,

> one can provide a special predicate, 'susceptible to observation' say, or 'causally efficacious', or, and so on, and recognize the ontological commitments of a discourse to be solely those objects falling under the extension of *that* predicate, to treat only *those* objects as existing (or *real*). (Azzouni 1998, p. 3)

The question Azzouni then asks is, 'How might we go about deciding between Quine's CRD and an alternative which takes causal efficacy, for example, as criterial for ontological commitment?' Azzouni considers several possibilities, but concludes in each of his cases that what one takes a particular argument as showing will depend on which CRD one is initially attracted to.[3] Thus, for example, if one argues that the ontological commitments of a discourse are to be determined by what it would take for the claims of that discourse to be true, then we need a CRD in advance to decide what would have to be the case for those claims to be justified. Similarly, if we attempt to invoke a metaphysical principle such as Occam's razor in arguing for the relatively conservative Quinean CRD over some alternative CRD which commits us to the existence of objects above and beyond the minimal set of variables quantified over in our best science, then we are in no better a position, since such a principle, which counsels us, other things being equal, to accept whatever theory commits us to the minimal amount of objects, can only be applied to adjudicate between theories after we have chosen a CRD. Thus, Azzouni concludes that,

> arguments supporting one or another CRD seem to either beg the question against the opponent, or be intelligible to begin with only if that CRD is already in place. The natural conclusion is that there is no bedrock below one's CRD, no place to get a foothold to apply pressure against an opponent. (Azzouni 1998, p. 10)

Since we cannot find any arguments that would pick out one CRD over any other, then even if we accept Quine's CWE (the claim that it is science that tells us what there is), once more we are left in a position where we have no hope of discovering the ontological commitments of our theories.

> So it looks like we must conclude that not only is it philosophically indeterminate what CRD is suitable, but that the question of what there is, understood in its philosophically broadest sense, is equally philosophically indeterminate. (Azzouni 1998, p. 11)

---

[3]This rough sketch skips over many of the details of Azzouni's particular arguments. For reasons of space I am unable to take a closer look at the details of each possibility Azzouni considers.

Once more, the conclusion is that the Quinean programme of naturalistic ontology must fail.

6.4.3 Neo-Carnapian Skepticism

We introduced the move to investigate alternative CRDs in light of our recognition of an apparent tension in Quine's naturalism. Quine's naturalistic call to look to science to tell us what there is (Quine's CWE) appears in tension with his CRD, since scientists themselves don't appear always to take their quantifier commitments as indicative of their ontological commitments in using a given theory. The upshot was that observations concerning the attitude of scientists to their quantifier commitments seemed to offer some hope for naturalistically inclined nominalists. It was suggested that we could agree with Quine that it is science that tells us what there is, but disagree regarding the question of how to interpret the ontological commitments of our best theories, thus opening the way to taking a differential attitude between the quantifier-commitments of our theories (e.g., by proposing realism only with respect to the quantifier-commitments of those components of our theories that we have good reasons to take as literal, or only with respect to those objects that play, e.g., a causal role). The arguments of Yablo and Azzouni, which speak as much against these alternative CRDs as they do against Quine's own identification of ontological commitment with quantifier commitment, are thus as much a blow to the nominalist ontological project as they are to that of the Quinean. There is a very real sense in which Yablo and Azzouni have both led us back to a Carnapian sceptical outlook, according to which *no* philosophical questions regarding ontology are legitimate. While Carnap's own arguments against the project of ontology are themselves not enough to show that project to be doomed, the arguments of both Yablo and Azzouni purport to show nevertheless that Carnap's suspicions were well founded. As Azzouni puts it, "Carnap's impression that there is something trivial about ontological commitment was deeper than the tools he used to support that claim." (Azzouni 1998, p. 14).

6.5 A More Thoroughgoing Naturalism

Can the post-Quinean ontological project be salvaged from Yablo and Azzouni's sceptical attacks? On the face of it, I will argue, it can. Yablo's and Azzouni's arguments, while effective against a particular conception of the Quinean ontological project (as involving applying a particular criterion of ontological commitment to our scientific theories), are not enough to thwart a more thoroughgoing naturalist approach to ontology, which would involve looking to science itself for an account of ontological commitment.

Consider again Azzouni's attack on Quine's ontological project, which argues that, even if we accept the Quinean CWE (that it is science that tells us what there is), we cannot find arguments to adjudicate between alternative ways of interpreting the ontological commitments of our theories. We cannot, as a result, find *a priori* reasons for favouring Quine's CRD over a variety of alternatives. Put this way, though, as an argument against the naturalist approach to ontology, Azzouni's attack loses much of its bite. For surely the purest of naturalists would *agree* that there is no use looking for *a priori* arguments for any particular philosophical criterion for recognizing the ontological commitments of a discourse. Rather than approaching our scientific theories with philosophical preconceptions about how to understand their ontological commitments, a philosopher who is serious about naturalism should look to scientific practices to discover what scientists take as requiring commitment to the existence a particular type of object. That we are unable to decide on a CRD in advance of looking at particular scientific practices should be of little concern for the naturalist who thinks that we should seek to discover what scientists themselves see as committing them to an ontology, rather than approach science with our own preconceptions about ontological commitment.

Azzouni's attack would remain potent if he could show that there is no way of discovering what scientists themselves take as evidence for the real existence of the objects posited by their theories – for example, if he could show that scientific practices fail to decide between an interpretation which sees scientists as taking quantifier commitment to signify ontological commitment and an interpretation which sees scientists as applying a special existence predicate to only some of those objects quantified over in those theories. But it is here where the evidence gathered in Maddy's discussions of scientific practice kicks in – just by looking at scientific practices it appears that scientists do make a distinction between those parts of their theories that they take to be really true as opposed to merely useful, and those objects posited by their theories that they have shown to really exist and those that they do not think of as really existing. Indeed, that there is a distinction at work between the objects that our theories are quantifier-committed to and the objects to which scientists take themselves to be genuinely ontologically committed to when they adopt those theories can be seen at a basic level in the difference between assertions that 'there exist *F*s' and 'there *really* exist *F*s'.[4]

What about Yablo's concern, that once we allow that some of our theoretical frameworks are adopted for their merely figurative, rather than literal, content, we will be unable to carry out Quine's ontological project because, in the controversial cases, we will be unable to separate literal from merely figurative uses of the

[4] I am grateful to Penelope Maddy for pointing me towards this way of escaping Azzouni's attack. I should also note that, following his 1998 paper, Azzouni himself came to the view that a more thoroughgoing naturalism could be used to escape his earlier scepticism (he indicated as much to me in conversation in 2002). Indeed, such an approach stands behind Azzouni's later defence of nominalism in his 2004 book *Deflating Existential Consequence*.

existential quantifier? Does this mean that, at least in the harder cases, in trying to read ontological commitments off scientists' usage, we will be unable to determine whether scientists mean their existentially quantified utterances to taken literally (i.e., as genuinely ontologically committing), or merely figuratively?

Yablo's worries about the possibility of determining whether theoretical claims are to be taken literally or figuratively stem from a concern that there may be no fact of the matter about whether scientists *intend* a literal or figurative interpretation. As Yablo (1998: 257–8) puts it, we often make utterances in a 'make-the-most-of-it spirit'. Thus,

> I want to be understood as meaning what I literally *say* if my statement is literally true—count me a player of the 'null game', if you like—and meaning whatever my statement projects onto via the right sort of 'non-null' game if my statement is literally false. It is thus indeterminate from my point of view whether I am advancing *S*'s literal content or not.

If the more thoroughgoing naturalist approach to ontology I have been suggesting, involving looking to science itself to determine a criterion of ontological commitment, requires us to determine how scientists *intend* their utterances to be interpreted, then the indeterminacy that Yablo identifies would indeed result in ontological scepticism.

However, as I argue in my (2010) book, the naturalist proposal to look to science to determine where to draw the line between mere quantifier commitments and genuine ontological commitments need not, and ought not, be understood as requiring the *hermeneutic* project of working out in what spirit scientists intend their utterances to be taken. Rather, I argue, our interest should be in uncovering which genuinely ontological claims are justified given scientific standards of evidence. "And," I argue,

> The question of which theoretical hypotheses are so justified may be quite independent of the question of the spirit in which our theoretical hypotheses area originally intended. (Leng 2010, p. 144)

Answering this question involves, I argue, inquiry – internal inquiry making use of ordinary scientific standards of evidence – into the role played by theoretical hypotheses in our theoretical successes. If our best explanation of the success of a theoretical posit requires the assumption that the object posited really exists, then this will support ontological commitment to that object. On the other hand, if we can explain the success of a theory that quantifies over *F*s from a perspective that sees *F*s as merely useful fictions, then the empirical success of our theory does not constitute evidence for *F*s. Applying this approach in *Mathematics and Reality*, I argue that it is possible to explain the uses we make of mathematical posits in empirical science from a fictionalist perspective according to which there are no mathematical objects. As such, I argue, ordinary scientific inquiry into the role played by mathematical posits in our scientific theories does not vindicate realism about the objects posited.

6.6 A More Thoroughgoing Skepticism?

The project of looking to science to discover what there is thus looks feasible even in light of recognition of a more complex relationship between the ontological commitments of scientists and the quantifier commitments of their theories. We can apply ordinary scientific standards of explanation to consider whether the best explanation of the successful use of a theory that posits Fs requires us to accept the real existence of Fs as opposed to taking Fs as merely useful theoretical fictions. However, recent work by Penelope Maddy has raised the possibility that this question, of whether, according to the concept of existence in play in natural science, we should conclude that there really exist Fs or that Fs are merely useful fictions, is itself ultimately unanswerable. According to Maddy, in hard cases (such as the case of mathematics) the question of whether, according to ordinary scientific standards of evidence, there really are Fs is unanswerable not because we cannot know in the cases that matter to us which side of the real/fictional divide the posits we are interested in fall, but because, for some values of our use of 'exists' in science fails to determine an answer to the question of whether Fs exist. Insofar as we do offer an answer to this question (positive or negative), our answer is best considered, in Carnap's terms, "a matter of decision rather than assertion" (Carnap 1950).

When we try to answer questions about what exists, clearly part of what determines whether an answer is correct or not is what we mean by 'exists'. Indeed, in the mathematical case, mathematicians are often frustrated by philosophical scruples about the existence of numbers since there is a very clear internal-to-mathematics sense of 'exists' according to which it's just obvious that numbers exist. If, when asking the question 'Do numbers exist?', we mean by 'exists', something like 'exists according to standard mathematics', then the question is answered trivially in the affirmative (as of course was recognized by Carnap). On the other hand, if, when we ask the question 'Do numbers exist?' we mean by 'exists' something like 'has spatiotemporal location', then at least most will agree that the question we are now asking can be answered in the negative. The naturalistic ontological project we have been discussing proposes that the sense of 'exists' that matters is the sense of 'exists' that is at work in natural science, so that when we ask the philosophical question of ontology we are asking what we have reason to believe exists according to our scientific understanding of this concept, not according to the internal mathematical sense of 'exist', or according to some metaphysical notion arising out of a prejudice in favour of spatiotemporal entities.[5]

[5]Why look to natural science for the appropriate sense of 'exist', rather than, as Maddy (1997) suggests, letting the internal standards for mathematics decide the question of whether there exist mathematical objects? At least two issues arise that speak against this proposal. First is what Gideon Rosen (1999, p. 471) calls the 'Authority Problem for Naturalized Epistemology', the problem of determining which internal claims of which frameworks to take seriously (if the internal standards for mathematics are authoritative in answering the question of whether there exist numbers, should the internal standards for, e.g., Christian theology be considered authoritative in answering the question of whether there is a unique God?). Second is a concern about univocality. If

Maddy's concern with this question is that, while our ordinary scientific uses of 'exists' can answer some ontological questions (we can rule out unicorns and rule in horses, for example), in the hard cases that matter to us our ordinary use of the concept simply does not determine a correct way to go on. To motivate this idea Maddy uses an example of Mark Wilson's (2006) of the limitations of the empirical concept of 'ice'. Ice is frozen water, but does this definition suffice to allow us to say in all cases whether something is to count as ice? According to Wilson (2006, p. 55), water is "a notoriously eccentric substance, capable of forming into a wide range of peculiar structures" (quoted in Maddy 2011, p. 106). There are, Wilson explains, in fact many ways that water can solidify, forming structures that do not neatly fit our paradigmatic examples of ice (e.g., as 'amorphous ice', solid but lacking the usual crystalline structure of the paradigm cases). When faced with these structures, ought we to consider them as ice, or merely ice-like? Wilson's point, which Maddy takes up, is that once we have described the structures and said in what sense they are, and in what sense they are not, like the paradigmatic examples of ice, at that point the question of whether to call these new structures 'ice' is not answerable to any substantial objective facts about ice, but simply depends on an arbitrary choice we can make about whether or not to refine our ordinary concept so as to apply it to these too. Either way we come to an acceptable way of describing the situation—either as a discovery of a new kind of ice, or a discovery that water can freeze solid without forming ice.

The case of the scientific concept of 'exists' is, in Maddy's view, similar. We have paradigm cases in science where we are happy to say of purported objects that they fall either side of the exists / doesn't exist divide. But these paradigm cases do not help us when it comes to purported abstract objects – and particularly mathematical objects. For Maddy, the question of whether a scientific look at mathematics and its role in science confirms the existence of mathematical objects is problematic because there are features of mathematical posits that display important similarities with the more paradigmatic cases of objects that we are happy to say exist (for example, the objectivity found within mathematics), and features that display important differences (the obvious case being lack of spatiotemporal location, but also, for example, lack of particular kinds of theoretical roles). The question of whether numbers really exist (or are mere fictions) may then turn out to be just like the question of whether certain solid water structures are really ice (or merely ice-like), answerable not by reflection on any substantial objective facts about what there is, but only by making a choice about how to refine our ordinary concept of 'exists'. Thus, Maddy (2011, p. 112) tells us,

'*Fs* exist' (in number theory) means '$\exists x Fx$' follows logically from the Peano Axioms, then we are tempted to ask the further question, but do we have reason to believe those axioms, and therefore that *F*s *really* exist. Carnap rejected such questions as illegitimate external questions. Quine's naturalistic move makes space for these further questions only by privileging the framework of natural science as the "ultimate arbiter of truth and existence" (Resnik 1995, 166).

> Our central questions—is pure mathematics of-a-piece with physics, astronomy, psychology, and the rest? is it a body of truths? do its methods confirm its claims?—these questions have no more determinate answers than 'is amorphous ice really ice?' ... [O]nce we understand how pure mathematics developed, how it now differs from empirical sciences, once we understand the many ways in which it remains intertwined with those sciences, how its methods work and what they are designed to track—once we understand all these things, what else do we need to know? Or better, what else is there to know?

If Maddy is right about this, then the question of whether we should say that (according to the notion of 'exists' at work in the natural sciences) numbers exist or are mere fictions has no deep answer. What we can say is that numbers as theoretical posits have some features of objects that we do take to exist, and lack other such features. What is distinctive about mathematics, for Maddy, is that the correctness of mathematical claims is grounded in objective facts concerning mathematical depth. Whether this objectivity should be taken to ground a claim concerning the real existence of mathematical objects (as Maddy's 'Thin Realist' claims), or whether Maddy's 'Arealist' is right in taking it that the objectivity of mathematics does not support the existence of mathematical objects, is not something that is ruled in or out by our scientific uses of 'exists':

> The application of 'true' and 'exists' to the case of pure mathematics isn't forced upon us—as it would be if Thin Realism were right and Arealism wrong—nor is it forbidden—as it would be if Arealism were right and Thin Realism wrong. Rather, the two idioms are equally well-supported by precisely the same objective reality: those facts of mathematical depth. These facts are what matter, what make pure mathematics the distinctive discipline that it is, and that discipline is equally well described as the Thin Realist does or as the Arealist does. Once we see this, we can feel free to employ either mode of expression, as we choose—even to move back and forth between them at will.
>
> The proposal, then, comes to this: Thin Realism and Arealism are equally accurate, second-philosophical descriptions of the nature of pure mathematics. They are alternative ways of expressing the very same account of the objective facts that underlie mathematical practice. (Maddy 2011, p. 112)

Is Maddy right about this, and if so, what does this mean for the Quinean project of ontology and our question of whether there is a fact of the matter concerning the existence of abstract objects. Two points: first, Maddy's arguments are aimed at showing that the empirical evidence together with our empirical concepts does not and cannot decide between the positions she is calling 'Thin Realism' and 'Arealism', which agree that the 'goodness' of mathematics is grounded in the objective depth of mathematical concepts and theories, but disagree about whether mathematics that is 'good' in this sense should also be counted as true. Maddy is certainly not saying that the empirical evidence can't rule out some pictures of mathematical objects as abstract. Thus, Maddy contrasts both Thin Realism and Arealism with a more traditional Platonist picture she calls 'Robust Realism', according to which there are mathematical objects and it is at least possible that an objectively fruitful, mathematically deep theory may get things quite wrong about these independently existing mathematical objects (so that the facts about what mathematical objects there are not grounded in facts about mathematical depth). Maddy's Thin Realist and Arealist will agree that nothing in either pure mathematical practice or in the use we make of mathematics in empirical science supports this kind of Robust Realist picture of mathematical objects, so to that extent a certain ontological picture is ruled out by our scientific notion of 'exists', even if some ontological questions remain unanswered. Second, and related to this, is the point that Maddy's 'no fact of the matter' position requires there to be substantial agreement on both sides on questions as to what matters in pure mathematics (e.g. in the search for new axioms for set theory), and on how mathematics gets applied in empirical science (both sides claim that a picture of mathematical correctness that focuses on the objectivity of mathematical concepts/theories can account for the success of mathematics in applications). If either accounting for pure or applied mathematics require us to posit an active role played by mathematical objects *as conceived by the Robust Realist*, then this might be enough to place mathematical objects unambiguously on the side of those things to which our scientific notion of 'exists' should be applied. As such, it would seem that the contemporary debate in the philosophy of mathematics over whether mathematical objects play a genuine explanatory role is exactly where the debate over the existence of abstract mathematical objects should be: if realists can make the case for mathematical posits playing the same kind of explanatory role as other theoretical posits, this would seem to speak against both Thin Realism and Arealism. On the other hand, if mathematical posits do not play this kind of substantial explanatory role then either Thin Realism or Arealism would seem to be compatible with the facts of mathematical and scientific practice. Only then does the question of the existence of mathematical objects become a matter of choice about how to extend the concept 'exists'.

6.7 Summary

Quine blocked Carnapian ontological scepticism by arguing that, since the theoretical framework we in fact use in empirical science has empirical confirmation (as a whole), we need not think of existence questions as making sense only internal to a theoretical framework. While it is true that individual frameworks might include rules that determine the answer *within that framework* to existence questions, this does not make the question of what *really* exists entirely framework relative. Rather, the question of what *really* exists becomes a matter of which existence claims within which theoretical frameworks are confirmed by the successful use of that theoretical framework in describing, predicting, and explaining our experiences. For Quine this means that existence questions are answered by looking at our best scientific theories and adopting as our ontological commitments the quantifier commitments of those theories.

Attention to the complexities of scientific confirmation led us to question whether Quine was right in thinking that, if we look to our best confirmed scientific theories to discover what we ought to believe that there is, we should automatically take the existentially quantified utterances used to express those theories as indicative of ontological commitment. If our theories include elements that are acknowledged by scientists to be merely useful instruments, then it would be strange to think that the successful use of those theories requires us to believe in the literal truth of components that scientists take themselves to have good reasons to treat instrumentally. Recognising this, we wondered whether we should adapt Quine's account of ontological commitment to recognise the fact that scientists themselves don't appear to take all of their theoretical components to be equally confirmed by their theoretical successes. Both Yablo and Azzouni have offered considerations that suggest that, once we recognise that our ontological commitments in utilizing scientific theories need not track those theories' quantifier commitments, the wheels come off the naturalistic ontological project, as there is no way of determining what we should take as indicative of genuine commitment to existence in science. I suggested a more thoroughgoing naturalism could respond to Yablo's and Azzouni's worries, by looking to scientific explanatory practices and looking to explanations of the roles played by various theoretical posits in our successful theories to discover whether the use of those posits supports realism about the objects posited. This was the project of my 2010 book, which focussed on the question of whether our account of the use made by mathematical posits in empirical science requires a realist understanding of the objects posited.

This more thoroughgoing naturalistic project has difficulties, though, if it turns out that, as Maddy (2011) suggests, the notion of 'existence' at work in empirical science is not determinate enough to answer questions of whether a given explanation of the role played by a given kind of theoretical posit counts as evidence for the existence of the objects posited. If Maddy is right about the notion of 'exists' we get from scientific uses of the word, the question of whether abstract mathematical objects exist or not might come down to a choice about how to continue to apply our term, rather than a determination of a genuine fact. I have suggested that Maddy may be right about this, but that this does not undermine the thought that there are *some* genuine facts of the matter about what there is. While Maddy argues that there may be nothing in our use of 'exists' that enables us to choose between Thin Realism and Arealism as two pictures of the nature of mathematics and the role of mathematics in science, nevertheless if Thin Realists and Arealists are right in their picture of the empirical role played by mathematics, both views stand together in denying a more robust Platonist account of the existence of abstract mathematical objects that sees these objects as being confirmed by their playing a more substantial empirical role. As such, even if the naturalist approach to ontology cannot answer the question of whether Thin Realism or Arealism is right, it may still allow us to reject a more substantial account of mathematical objects as abstract as offered by the Robust Realist. So there may be *some* facts of the matter about the existence of abstract objects (e.g. that we have no reason to believe that they exist in the robust sense), even if on a thinner conception of mathematical realism the realism/fictionalism question cannot be answered.

References

Azzouni, J. (1998). 'On 'On what there is''. *Pacific Philosophical Quarterly, 79*, 1–18.
Bird, G. H. (1995). Carnap and Quine: Internal and external questions. *Erkenntnis, 42*, 41–64.
Carnap, R. (1950). Empiricism, semantics and ontology. *Revue Internationale de Philosophie, 4*, 20–40.
Leng, M. (2010). *Mathematics and reality*. Oxford: OUP.
Maddy, P. (1992). Indispensability and practice. *The Journal of Philosophy, 89*, 275–289.
Maddy, P. (1997). *Naturalism in mathematics*. Oxford: Clarendon Press.
Maddy, P. (2011). *Defending the axioms: On the philosophical foundations of set theory*. Oxford: OUP.
Quine, W. V. (1948). On what there is. In *Review of metaphysics*. Reprinted in Quine (1961, Vol. 2, pp. 1–19).
Quine, W. V. (1951). On Carnap's views on ontology. In *Philosophical studies*. Reprinted in Quine (1976, pp. 203–211).
Quine, W. V. (1961). *From a logical point of view* (2nd ed.). Harvard: HUP.
Quine, W. V. (1976). *The ways of paradox and other essays* (2nd ed.). Harvard: HUP.
Resnik, M. D. (1995). Scientific vs. mathematical realism: The indispensability argument. *Philosophia Mathematica, 3*, 166–174.
Rosen, G. (1999). Review of naturalism in mathematics by Penelope Maddy. *British Journal for the Philosophy of Science, 50*(3), 467–474.
Wilson, M. (2006). *Wandering significance*. Oxford: OUP.
Yablo, S. (1998). Does ontology rest on a mistake? *Aristotelian Society, 72*, 229–261.

Part III
Fictionalism or Realism in Philosophy of Empirical Sciences

Chapter 7
An Ensemble-Plus-Standing-For Account of Scientific Representation: No Need for (Unnecessary) Abstract Objects

José A. Díez

Abstract In recent discussions on modelling and scientific representation, it has been argued that, in order to account for some key features of the practice of modelling in science, mainly the existence of unsuccessful representations and also of successful yet inaccurate or idealized ones, it is necessary to accept abstract objects, or their fictional versions. The goal of this paper is to present a new account of scientific modelling and argue that according to such account there is no need of abstract objects in a strong sense.

Keywords Scientific model · Scientific representation · Structuralism · Functionalism

7.1 Introduction

In recent discussions on modelling and scientific representation, it has been argued that, in order to account for some key features of the practice of modelling in science, mainly the existence of unsuccessful representations and also of successful yet inaccurate or idealized ones, it is necessary to accept abstract objects, or their fictional versions. The goal of this paper is to present a new account of scientific modelling and argue that according to such account there is no need of abstract objects in a strong sense. Limitations of space do not allow us to do a detailed

Research for this work has been supported by the research project FFI2016-76799-P, Spanish Ministry of Science, Innovation and Universities. I want to thank José L. Falguera, Roman Frigg, James Nguyen, Pablo Lorenzano and Stathis Psillos for comments and criticisms to earlier versions of this paper.

J. A. Díez (✉)
LOGOS Research Group, Barcelona Institute of Analytic Philosophy, University of Barcelona, Barcelona, Spain
e-mail: jose.diez@ub.edu

J. L. Falguera, C. Martínez-Vidal (eds.), *Abstract Objects*, Synthese Library 422,
https://doi.org/10.1007/978-3-030-38242-1_7

justification of the account and its assessment vis a vis its rivals (tasks that will be done in a different paper), but the brief presentation will be enough to make clear the sense in which it may dispense with (unnecessary) *abstracta*.

According to the Ensemble-Plus-Standing-For (EPS) account, representational modelling is the scientific practice that consists in building ensembles and taking some elements/parts/features of the ensemble as standing for other entities. What matters for making this general idea both precise and plausible is of course its details. Two main aspects on which the account crucially depends are, first, the neat distinction between the existence/performance and the adequacy/success of the representation (this distinction is present in several authors, e.g. Callender and Cohen 2006 talks of constitutive vs normative features, Frigg and Nguyen 2017a, b of distinguish between analyzing the notion of epistemic representation and analyzing the in/accuracy of such a representation); and second, the introduction of some essential pragmatic elements, crucially the idea of logical-form-in-use (together with other more standard context-dependent features). I first introduce the three main problems of scientific representation that the account aims to answer, and then clarify the many parameters implicit/explicit in representational sentences. Second, I summarize the EPS account and make some necessary clarifications about some of its clauses. Finally, I defend that on such an account the *strong* reference to abstract objects is dispensable. If successful, the dispensability of abstract objects amounts to an additional virtue of EPS vis a vis its "abstractivist" competitors.

Before starting with the three problems, two caveats. First, the account aims at being both enough monistic and enough substantive. This means that, although the times for providing and strictly sufficient conditions have passed, the analysis aims to give necessary conditions that are "nearly sufficient", i.e. as strong as possible (which implies that given two alternative analyses, the strongest is, ceteris paribus, the best). The goal is then to exclude all cases of non-modelling, and include all *paradigmatic* case of modelling; and to do so not by mentioning minimalist platitudes but by providing substantive conditions that explicate the concept. The account tries to resist both pluralist and minimalist withdrawals that proliferate nowadays.

Second, the account is confined to the *representational use* of scientific models. Scientific models can be used to serve other goals. For instance, they may be designed just to test whether certain empirical hypothesis or theory is coherent without any attempt to tell how reality actually is (e.g. Tobin 1970's ultra-Keynesian model). Other models (e.g. some computer simulations, Humphreys 2014; Weisberg 2013; Winsberg 2010) may have a *merely* predictive function, that is constructed to carry out relevant observational predictions in a given field but without the aim the non-observational components representing anything actually present in the phenomenon (of course, merely predictive models are representational in a minimal sense, namely they represent the observable phenomena they predict). Some models are only *partially merely* predictive, i.e. with elements that aim to correspond to observable predictable phenomena, some of other components that play a role in the model and aim to correspond to something else in the phenomena, but still other components that also play an essential role yet without any representational import (e.g. the Burridge and Knopoff 1967 block model of earthquakes; or the ancient Antikythera Mechanism for the solar system, cf. Carman et al. 2012; Jones 2017). And there may be other non-representational uses, for instance, as heuristic tools such as Maxwell's vortex fluid dynamical model of the causal relation of electricity and magnetism (Morrison 2015), or the three-sex model reproduction in population dynamics (Weisberg 2013); or some of the uses of models as mediators (cf. e.g. Morgan and Morrison 1999), or as epistemic artefacts (Knuuttila 2011), or as explanatory fictions (Bokulich 2009), or as tools for theory construction (Hartmann 1995); all these uses may be (partially) non-representational as well (cf. Peschard 2011 for not-only-representational uses of models). On the other side, not only scientific models *stricto senso* represent, other scientific objects (theories, engineering graphs, scientific images/drawings, geographical maps, ...) also represent. The analysis aims to apply to any scientific *representans,* so I will be using 'scientific model' in a wide sense as referring to any scientific construct with representational goals, scientific theories included. As for non-scientific representations, although the analysis might have some application to some of them, it does not aim to apply to non-scientific representations in general and must then be taken as confined to scientific cases.

After these caveats, let us just list some paradigmatic examples of scientific *representantia* that, without any aim of exhaustion, are representative of the main kinds of modelling one can find in scientific practice that our analysis aims to apply to: the double-helix, and the triple-helix models of DNA; the ideal pendulum model; the kinetic theory of gases and its billiard ball analogy; the phlogiston and oxygen theories of combustion; Ptolemy's, Brahe's and Copernicus' models of the Solar system; Newton's and Einstein's theories of gravitation; the Lotka-Volterra predator-prey model; the above mentioned Burridge-Knopoff block model of earthquakes; the Bak-Sneppen model of species replacement; the computational model of the mind; Thomson's, Rutherford's and Bohr's models of the atom; XIX century corpuscular and wave theories of light; the pipes-reservoirs-fluid Philips-Newlyn hydraulic model of economic dynamics; the numerical representation of magnitudes in Representational Measurement Theories; the scale Mississippi River Basin model; the scale aircraft wind-tunnel simulation; the use of Drosophila as a model of general genetic phenomena; the experiments with mice for modelling human pain or reactions to drugs; the Mercator and Gall-Peters two dimensional maps of continental land. This is only a small sample of the vast plethora of models and representations in scientific practice. They cover a whole variety of scientific models/representations: material models; theoretical models; qualitative models; mathematical models; scale models; phenomenological models; analogical models; idealized models; organism models; in vivo and in vitro models; reductive models; etc. The goal of the EPS account is to resist the current pluralistic attitude and present a sufficiently unified and substantive account that applies to this plethora of scientific models explicating what they all have in common.

EPS aims to answer the following three key problems of scientific representation:

(PD) Problem of directionality: *In virtue of what does the direction of the representation run from the model (M) to the target (T)?*
(PP) Problem of existence/performance: *In virtue of what does M represent T?*
(PA) Problem of success/adequacy: *In virtue of what, if M represents T, does M do so successfully/adequately?*

There usually is other main problem mentioned (see e.g. Frigg and Nguyen 2017a, b; Gelfert 2017), namely the problem of ontology (PO): *What kind of entities are models?* EPS does not have a particular answer for PO, but constraints the kind of answers that one might give. It is in this regard that the account has consequences for some answers to PO in terms of abstract entities that will be discussed in the last section.

With respect to these three problems, the leading idea is to treat scientific representation as analogous to speech acts: in speech acts, we do things (asserting, inquiring, commanding, ...) with words; in (representational) modelling, we also do things, sometimes with words, but on other occasions with drawings, figures, symbols, material pieces or other items. The most important consequence of this analogy is the neat distinction between Performance conditions (P-conditions) and Success conditions (S-conditions). A speech act may fail even in being performed: one may intend to perform it and fail if its performance conditions have not been met. Once the speech act has been properly performed, it may fail in a second sense, namely not generating the expected action (forming a belief, obeying the command, answering the requested information, etc.) in the audience. The same applies to modelling/representing. For the representation to be performed, certain conditions must be met (problems of directionality and existence/performance above). And once the representation has been properly performed, additional conditions must be satisfied for the representation to be successful in the second sense (problem of success/adequacy above). Models/representations also have certain goals constitutively associated to them, and the model succeeds in this second sense if its constitutive goal is satisfactorily reached. If one accepts this,[1] then this part of the analysis has to come in two subsequent steps. First, one has to provide an answer to PP, and then to PA (including as precondition that the performance conditions have already been satisfied).

In order to avoid terminological confusions, let me clarify the use of some crucial expressions:

– First, I will use ‘model-description’ to refer to the linguistic items (if any) used by the modeler to describe the model or to instruct its material construction (if the model is material): ink-words in papers and text-books, lists of instructions,

[1] One might not, for instance defending that no-representation simply is a limiting case of misrepresentation, that is that mere will suffices for representing and if one does it completely badly then we can talk of no-representation. If I understand them well, this is the stance taken by Chakravartty (2009) and Ducheyne (2008), for whom some degree of success is necessary for the existence of representation.

written equations, etc. (I say "if any" because some models may come without physical signals/symbols but directly described-instructed "by the mind").

– I will use 'model' to refer to what the model description describes or instructs to construct, which according to EPS is basically an *ensemble* of entities together with a *standing for* relation. There is more to be said below but by now it suffices to say that, for instance, in material models such as the physically constructed double helix DNA model, the ensemble is the complex physical entity constructed. Or in a mathematical model the ensemble (or one subpart of it) is a certain mathematical function with some properties such as being over the real numbers, being differentiable, etc.
– I will use 'target system' to refer to the phenomenon in nature that the model aims to represent/account for. Here we have to take a decision. One may individuate the target as (allegedly) constituted by the entities the model attributes to it. If one goes this way, then when a model attributes to the target an entity that actually does not exists (phlogiston, caloric, ether and other similar cases), then there is no target, the model is targetless (accepted e.g. by Frigg and Nguyen 2017a, b; notice that, thus, strictly speaking, the model does not attribute the non-existent entity "to the target", since there is no target either). The main problem with this option is that this forces us to what I take to be a counterintuitive talk: on the most common talk both phlogiston and oxygen models have the same target, they both represent (differently) combustion-calcination; the same in other cases (caloric and kinetic theories have the same target, they both represent heat phenomena; ether and electromagnetic models have the same target, they both represent light phenomena; etc.). I think this is the most common talk that is worth to preserve and that the alternative option cannot naturally[2] preserve. I will then use 'target' to refer to the *previously* identified and individuated natural phenomenon (by, say, features t1, .., tn), towards which the model is addressed;[3] for instance, though phlogiston model attributes phlogiston to the target, and oxygen model attribute oxygen to the same target, neither phlogiston nor oxygen *individuate* the target (they may, if the model turns out to be correct, "be in the target", but the target is not individuated by them). This does not rule out the possibility of models without targets, Maxwell's models of the ether or n(>2)-sex models in population dynamics are cases in point, although, as said above, I take them as being heuristic rather than representational strictly understood.

[2]One might preserve this talk making the target be the phenomenon plus what "the true model" attributes to the world behind the phenomenon; the target would be the content of "the correct model", whichever it is. I think this option is not "natural" since, given the uncertainty about truth, we would never know what the targets are, which I find counterintuitive. The only thing we could say with certainty is that the target is the phenomenon "together with whatever is behind it", which I do not think is very illuminating, nor that it provides any benefit that my own, simpler talk does not.

[3]It follows that as far as a target exists, its individuative features t1, . . . , tn also exist; how these features are themselves individuated (partially theoretically?, fully observationally?), is left here as an open issue.

– Finally, I will use 'model content' to talk of "what" the model expresses, the alleged state of affairs that according to the model exists in the world (for this it is crucial to bear in mind that not all the components of the models are intended to be projected to the world, more on this below). Note that I do not use 'refer' here, for it is one important feature of this account that when the model fails (e.g. phlogiston) there is no entity/thing in the world to which 'model content' refers. I'll discuss this in the last section.

We have been talking as if the representational relation involved only two parameters, model and target. Nevertheless, as many have emphasized (prominently Giere 2004, 2010, and van Fraassen 2008), this is a simplification: "M represents T" makes implicit reference to other parameters. In my account, at least four other additional parameters:

Subject S. Representations do not obtain *in abstracto*, they are the result of a scientific practice performed by particular individuals or communities. We use the variable S to refer to the relevant (individual or collective) subject that performs the representation, i.e. for the *representator*: "S uses M to represent T". As will become apparent later, the subject will be determinant in answering the problems of directionality and existence.

Respects R. Given a subject, a model and a target, the assessment of the model as successful or adequate often depends on certain respects, the features/properties of the model that are taken as relevant for the representation. As the bidimensional continental maps example above exemplifies, the same model may be adequate taking into consideration some of its respects or properties but not others. The Mercator model is adequate with respect the shape of the land, but not with respect the area dimension; while for the Gall-Peters the opposite is the case. It is then relative to some of its respects/properties that are taken into consideration, and not others, that a model is assessed: "S uses M to represent T in respects R (of M)".

Purposes P. Respects come with purposes. If the shape-respect of Mercator map is relevant for assessing it, it is so relative to certain purposes, in this case obtaining faithful information about a continent's real shape; and analogously for the Gall-Peters map and the purpose of obtaining faithful information about the dimensions of continental land areas. And, as it happens in this case of bidimensional continental maps, it may very well be that there is no model adequate for all purposes: If a bidemensional map correctly represents continental shape, it does not correctly represent area dimensions; and the other way around. "S uses M to represent T in respects R for purposes P".

Context C. Finally, the context of use C is determinant for the assessment. Representations are not all/nothing successful; success comes in degrees. Given the same subject, model, target, respects and purposes, a very stringent context may assess the model as unsuccessful if the degree or level of accuracy in which the relevant respects satisfy the purposes is below the demanding limit considered desirable by the context; while in other, less demanding context, the same model may be regarded as sufficiently adequate in the same respects for the same purposes. Context C is then essential in determining the degree of accuracy (and maybe other factors) for the assessment: "S uses M in C to represent T in respects R for purposes P".[4]

7.2 The EPS Account

With the above preparatory notions at hand, we can now present EPS. Since we separated the problems of existence and success, EPS comes in two steps. First, we explicate the existence/performance of representation (EPS-P) in terms of conditions that must be met for an agent that intends to perform a representation to actually perform one. In this regard, mere intention does not suffice. Intention suffices for answering PD, i.e. for determining the direction of the representation from M to T, but, as in speech acts, other conditions must be met for properly performing the act of representation. These conditions make use of the key notions of "ensemble" and "logical congruency constraint".

An ensemble is simply a bunch of entities (individuals, properties, relations, ...) articulated in a specific manner; for instance, two individuals exemplifying two first order properties and related by a first order binary relation; or three first order properties related by a second order triadic relation. Notice that for a bunch of entities to be able to form an ensemble they must have the appropriate logical categories, two individuals and a ternary first order relation, or one individual with a second order monadic property, cannot form ensembles. Whether there is a realm of "simple" entities with a logical-form-*in-se* (like in Wingenstein's *Tractatus*) is an open question, but EPS does not depend on it since in practice we always face entities that are not simples, and it is in the act of representing that *we take/use them* as-simples and attribute a logical category to them, what I call a logical-form-*in-use*. A red circle may be taken as simple, as one element of (the ensemble consisting in) the London tube map; or it may be taken as an ensemble itself (consisting in one individual instantiating two monadic properties) that, e.g., may represent that King Arthur is fat and honest. So, in representing the modeler takes some entities as simples, attribute certain logical category to each one and constructs an ensemble out of them.

The model is an ensemble plus a standing for relation. That is, an ensemble in which the subject takes some of its (taken-as-simple) parts as standing for other (usually different) entities. For the representation to be performed, the bunch of stood-for entities must themselves be able to form an ensemble, that is they must have the proper logical category (also *in-use*) to be "ensembleable" to each other like their originals in the model are. This is easily satisfied if the entities in M stand for entities of the same logical category. But this is too stringent. For

[4]Some might feel that an additional parameter is lacking, for as some have suggested (van Fraassen 2008; Elgin 2010), it is essential in representation that the target is *represented-as* something (for a summary of representing-as accounts, see Frigg and Nguyen 2017a, b). I do not include it for I think that the same relevant features may be captured with respects, purposes and context.

instance, there is nothing that conceptually excludes that the ensemble consisting of an individual instantiating a first order monadic property represents a second order dyadic relation instantiating a third order monadic property. If we relax the same-logical-category condition for this more flexible constraint, we obtain the *logical-congruency* constraint: the respects in M must stand for entities that "preserve their logical congruency/combinability". A precise definition is beyond our present goals, but this informal characterization suffices: it excludes e.g. aimed models consisting in an ensemble constituted by two individuals instantiating a binary first order relation, and aiming the two individuals stand for first order properties and the first order binary relation stands for a second order triadic relation; such ensemble with such standing for relation violates the congruency constraint and therefore the aimed representation is not even performed.

With these tools at hand we can now introduce the first step of our explication, the conditions for the *performance* of a representation:

(EPS-P) *S uses M to represent T in respects R with purposes P in context C:*

(i) *S has in C the intention-of-representing T (individuated by features t1, .., tn) with M in respects R, for purposes P;*
(ii) *in C, M is for S an ensemble of entities, those of which are respects are such that S intends, actually or fictionally (at least one respect actually), that they stand for other entities in T, so that the logical congruence constraint is satisfied;*
(iii) *in C, all target individuative features t1, . . . , tn of T are stood-for respects R of M.*

The intuitive idea is that for a representation to be performed, the subject must select a target phenomenon, build an ensemble (in which she attributes logical categories to constituents of model and target), be logically coherent (satisfy the logical congruency constraint) and materially coherent (aiming the stood for objects to belong to the previously selected target), and guaranteeing that all target individuative features are stood for respects in the model (although the agent may intend respects in M to stand for something not individuative of M, and fail, without this ruining the performance).

Some clarifications are in place. Note, first, that (in a Tractarian mood) representing "for the whole" is completely different from representing "for its parts": the relation of M representing T is of a different nature than the standing for relation of its parts/respects. Then, clause (i) explicates directionality: for the direction of the representation be from M to T the mere intention of S suffices.[5] Yet, as the hyphenation connotes, this S-intention-to-represent T is a primitive concept, not analyzable in terms of "intending" and "represent" on pain of circularity.

As announced, subject's intention, though necessary, is not sufficient for representation, two additional conditions must be met. Clause (ii) demands that M is taken by S in C as an ensemble (thus formed by taken-as-simple parts with some

[5]Note that, for this intention to be present, the target T must exist, the target may lack many of the features attributed by the modeler, but not all. So, according to our use, one cannot intend to represent no-thing. A "representation" of say, Atlantis, who is acknowledged as non-existing, is not a representation in strict sense (unless what one wants to represent, the target, is intended to be a fictional entity as fictional, or an abstract conceptual entity).

logical-form-in-use) and that S intends that the relevant components (respects) of M stand for other entities respecting the logical congruency constraint.[6] Here, "taken as simple" means that the inner structure, if any, of the component does not matter. In the material double helix model of DNA, the "nature" of the individual pieces joined in certain order and form, whether they are made e.g. of wood or metal, does not matter; what matters only is certain property they have, e.g. having one of four colors, and certain relations they enter, e.g. being paired with other piece, and being in certain spatial order. These components of the model are aimed to stand for elements of the target: the individual pieces for chemical bases, their color property for a chemical property (being Adenine, being Thymine, ...), their impairing relation for chemical bounds, and their order-shape relation by order-shape relation in a biochemical chain.

We face now a complication. We know that many times S may intend that some part of M stands for something that actually does not exist, e.g. phlogiston, caloric, and the like. This problem is solved by saying that *S intends that the respects stand for* (existing) entities instead of simply saying that the respects must stand for entities. Standing for is factive, if x stands for y both x and y exist. Then the condition cannot be that respects stand for other things since many times some of them do not stand for anything. That is why (ii) talks, not of standing for, but of S's intention that respects stand for other entities, and in cases like phlogiston this intention is present, though fails. In the last section we will see the consequences of this move for abstractionism. Note, tough, that the cases that oblige to introduce some (weak) kind of fictionalism are not these but those in which the model is partially representational and partially not (like the ones mentioned in Sect. 7.1), that is, those in which S essentially uses some relevant respects of M for obtaining the information that wants to transfer from M to T, *but* some of these respects are *not aimed* to have a correspondence, *not even upon idealizations*, in reality. S does not aim these non-projected respects to stand for anything (not even approximately), nevertheless she proceeds *as if* they were projected. They do not stand for; they are not even intended that they stand for; the model just proceeds *as if* these elements were aimed to be projected although it is explicit that they are not so aimed. In this sense its projection is just a *pretense*, a *fiction*.[7] Of course not every respect in M may be so intended, thus the condition that at least one does not.

---

[6] *Ensemble* is a conceptually (not metaphysically) primitive notion, only introduced by examples; it is not explicated in other terms such as set-theoretical structure or the like. This avoids the problems of set-theoretical structuralism in representation (Frigg 2003), but, admittedly, at the cost of taking it as primitive (for a discussion of models as representational structures, cf. Morrison 2008; Thomson-Jones 2011).

[7] Admittedly, the representational success of the whole model, based on the modelized behavior of this fictional element with the other elements that are really intended to stand for things in T, is somehow mysterious. I do not think this is a deficiency of this analysis, but rather the analytical reconstruction of how things are in these quite unusual cases of non-projected respects. Every analysis has to face these odd cases, in our case we do this appealing to intention of fictionally-standing-for.

The final clause (iii) accounts for the fact that we cannot represent (not even wrongly) if the standing for relation does not appropriately connect the respects in M and the individuative elements of the target towards which, according to (i), S intends M to be addressed. Suppose that I intend-to-represent the killing of Caesar. For doing that, I construct on top of my table an ensemble made of two medium size pieces of wood and a little one of plastic, taken as individuals, and certain spatio-temporal relations. But then, in my intended standing for relation, one of the wood pieces stands for Maradona, the other for the 1986 English goalkeeper, the plastic piece for the ball in the quarter-final, and the spatiotemporal relations on the table for spatiotemporal relations in the Mexican stadium (this aimed standing for relation satisfies the logical congruency constraint). I do not think that, *given that* I have intentionally addressed the model as targeting the killing of Caesar, we can say that I am nevertheless representing it, although very wrongly. If the objects that the representator intends to be stood-for do not belong to representator's intentional target, the representator is "objectually" incoherent and, even if logically congruent, there is no representation at all.

After the explication of the existence of representation, we are now in a position to give as a second step the analysis of the success or adequacy of a (n existing) representation:

(EPS-A) *S use of M to represent T in respects R with purposes P in context C is adequate:*

(i) *S uses M to represent T in respects R with purposes P in context C;*
(ii) *all entities that respects aim to stand for exist;*
(iii) *entities that respects aim to stand for, behave to each other in T C-like their corresponding respects in M;*
(iv) *purposes P are achieved, and are so in virtue of (iii) (respects, if any, that aim to stand for only fictionally, matter for demonstration in M but not for behavior alikeness in T).*

EPS-P combines the grain of truth of previous inherentist and functional accounts. I qualify as *inherentist* the accounts according to which the representation obtains in virtue of the inner natures/structures of model and target and of how they relate to each other. Although the proponents are unclear and many simply ignore the distinction between performance and adequacy of the representation, these accounts have some plausibility only as accounts of adequacy, not of performance. Inherentist accounts come in different versions, depending on the relation between M and T that explicates adequacy, from similarity (e.g. Giere 1988, 2004, 2010; Teller 2001; Weisberg 2013) to several kinds of morphisms or structure preservation (Suppes 1970, 1974, 1989; Sneed 1971; Balzer et al. 1987; Mundy 1986; Swoyer 1991; Bueno 1997; Bueno et al. 2002; Da Costa and French 2003; Bueno and French 2011; French 2014; Van Fraassen 1980, 2008; Contessa 2007). We cannot enter into the details here, but all these versions suffer from being either too weak (as similaritivism in general and some weak partial homomorphism versions) or too strong (as many morphism versions). In general, if the proposed relation is too weak, it does not exclude cases we want to exclude; and if it is too strong, it excludes cases we do not want to exclude. Nevertheless, the grain of truth is that *some* communality-of-behavior relation must obtain in every particular context for the representation being successful, yet *no particular* such relation is demanded in absolutely all contexts. The moral is that the context C specifies the particular communality-of-behavior relation relevant in C. This is what clause (iii) expresses.

With regard functionalisms, the main ingredient in their analysis of scientific representation has to do with the function that the model performs with respect to the target, some kind of *information transference* or *surrogative reasoning* (some representatives are Hughes 1997; Suárez 2004, 2010, 2015; Contessa 2007, 2011; Ducheyne 2012; and also, to some extent, some advocates of the representation-as account, e.g. Elgin 2010; van Fraassen 2008; Frigg and Nguyen 2017a, b; and some fictionalist accounts as well, such as Frigg's 2010). Although many authors do not distinguish explicitly between performance and success either, it is plausible to regard surrogative reasoning/information transfer as a condition for performance (since all accept that the information inferred may be wrong of the target), while the *correctness* of the drawn inferences as a condition for adequacy/success. EPS-P incorporates the functional conditions of performance in the fact that the satisfaction of EPS-P conditions habilitates for transferring (correctly or wrongly) information from M to T. And EPS-A(iv) incorporates the functional condition of adequacy, namely that the function/purpose is satisfied. Note, though, that this clause goes much further, or deeper, than regular functionalism, for it claims that the satisfaction of the purpose must be *in virtue of* there actually be a communality of behavior. Thus EPS-A does not combine structural and functional conditions by mere conjunction, but does it in the adequate, deeper manner, namely by showing that the satisfaction of the function *is grounded in* the communality of behavior.[8]

As for the other EPS-A clauses, (i) is simply the precondition of performance, and (ii) demands that the respects in M that S actually (not fictionally) intends to stand for entities in T, actually do so, thus that the intended stood for entities exist. This concludes the summarized presentation of EPS. Notice that once representational adequacy is explicated, misrepresentation is immediately explicated as inadequate representation: *S misrepresents T with M in respects R with purposes P in context C if and only if: (i) S uses M to represent T in respects R with purposes P in context C; and (ii) S does so unsuccessfully.* We then have three different sources of misrepresentation. The representation may fail due to the non-existence of a postulated entity (as postulated). Or all the postulated entities (together to features t1, . . . , tn of T) exist, but they do not behave to each other C-like the corresponding constituents of the model behave in the model. Or, finally (and admittedly unusual), the postulated entities exist, they behave C-like the constituents of the model in the model, but either (due to unexpected factors) the purposes are not reached, or they are reached but not in virtue of the communality of behavior.

[8] Some functionalist, e.g. Suárez (2015), also claim that the function must be performed in every case due to specific relation between M and T, but explicitly denies that this is part of the concept of representation.

7.3 A Tractarian Dispensation of (Unnecessary) Abstract Objects

EPS has a clear tractarian inspiration, but incorporates a conceptually necessary functional component and, importantly, also conceptually necessary strong pragmatic elements: there are no "simpleness" and logical categories in-se but only in-use in a given context, and it is context that determines the specific communality-of-behavior *and* the degree of satisfaction according to which the eventual success of the model must be assessed. This makes also room for a widespread phenomenon in scientific representation, namely inaccuracy, idealization and the like. One kind of inaccuracy, already mentioned, comes from cases in which some parts of M that are essential for gathering the information about M that is going to be transferred to T, are not intended to correspond to anything in T, not even approximately. We have seen that EPS-P(ii) accounts for it. The other, more widespread kind of inaccuracy is related to inexactness, distortion or idealization: perfect spheres, infinitesimal distances, frictionless planes, point-masses, infinite particle collections, perfectly isolated populations, and the like. Models include these individuals and properties or relations, but the modeler does not aim that the parts of the target for which they stand have exactly these features; it is aimed only that they *approximate them to a certain degree* determined by C: almost perfect spheres, very small distances, very small massive bodies, indefinitely large finite collections, almost isolated populations, etc. It is then the context that determines the degree of accuracy required for the purposes in point. This leaves a lot of work for the context, but this is how things are and there is nothing else we can do *at the conceptual level*.

EPS is sufficiently unified and substantive, and free of clear counterexamples, for counting as a plausible monistic account, and, so I claim, it fares better than its competitors, including Hughes' (1997, 2010) DDI account, Suárez's (2004, 2015) inferentialism, Bueno and French's (2011) partial structuralism, Contessa's (2007) interpretationalism and the recent (Frigg and Nguyen 2016, 2017a, b) DEKI account. In this last section I argue that EPS has an additional virtue, namely to make the (strong) postulation of abstract objects dispensable. I qualify "strong" since some kind of abstract objects are unavoidable for every party (but perhaps radical nominalists), and what is at stake is whether one can account for scientific modeling without postulating additional abstract entities. I take that it is a virtue of EPS that it can.

There effectively are some kind of abstract objects that are unavoidable, also for EPS. To start with theoretical, non-material, models, such as ideal pendulum or Lotka-Volterra, the ensemble that, together with the intended standing for relation, constitute M, has as constituents some abstract objects, e.g. a differential equation, or a point, or an "isolated" set.[9] Some fictionalists (e.g. Godfrey-Smith 2006a, b; Cartwright 2010; Frigg 2010) claim that these entities are parts of "fables" or

[9]Remember that the constituents of the (theoretical) models/ensembles are no the aimed stood for entities, i.e. the physical differentiable *velocity*, a point *particle*, or an isolated *populations*, but

imaginings launched by the text/prop, but it is not clear at all that this makes them non-existents (Frigg leaves the issue open and Godfrey-Smith remains silent about the issue). To some fictionalists (e.g. Thomasson 1999) fictions are or involve abstract entities. Others (e.g., Toon 2012; Levy 2015) want to defend that imagined situations and entities simply do not exist in any sense. For instance, that differential equations (numbers, points, etc.) do not exist but are just imagined, and particular imaginings/fictions and their parts do not really exist.[10]

As I said, radical nominalism shares this goal, but it is a highly controversial stance anyway, and it is not clear there are others less controversial. I do not think there is a way to explicate that two modelers are imagining the same differential equation or point mass without abstract objects, at least set-theoretical equivalent classes. In any event, if there is such a way, and it is considered ontologically worth, then nothing in EPS precludes to incorporate this fictionalist-non-existent ingredient in the account. As for material models, e.g. Watson and Crick or Phillips-Newlyn models, the constituents of the physical ensemble are in principle spatio-material individuals, properties and relations. Nevertheless, there may be different material replicas of the same model, and unless one decides that there are as many different models as material tokens, all are replicas of the same model, reconstructed at least as an equivalence class of material entities, which is an abstract object (the same goes for its parts).

Another source of abstraction is what we called "model description", a text describing the model. Again, although a particular text is a material entity, different texts may describe the same model and, in a relevant sense, they are the same description, which corresponds to a type or equivalence class. Similar considerations apply when the description is not in physical texts but in mental acts/thoughts.

I think that all accounts have to face these prima facie sources of abstractions in a similar way, whichever the best way is (including radical nominalism, if there is a plausible version). Different accounts need not differ with regard to them *as accounts of scientific representation*, so I will focus on other possible sources of abstraction that I call abstractionism in the strong sense, and which I claim EPS permits to dispense with. There are two such sources. One is general and corresponds to what in Sect. 7.1 I referred by "model content", that is "what" the ensamble-plus-the-intended-standing-for-relation represents, the "state of affairs" that M represents. When the model is correct/adequate (we leave inexactness aside for this discussion), there is a concrete state of affairs in the world which is "that". But: what when the model is incorrect? For instance, if I build an ensemble with two pieces of wood one left to the other, and intend that the piece on the left stands for Caesar, the one on the right for Brutus, and the relation *being-left-to* standing for the relation *stabbing*. There is no concrete state of affairs in the world such that Caesar

entities that are intended to stand for them, i.e. differentiable functions, mathematical points and sets (of somethings that are aimed to stand for biological individuals).

[10]Remember, again, that we are now talking of the constituents of the models, not of the entities attributed to the target.

stabs Brutus. Is there no model content then? Some, in a Meinongian fashion, may defend that in such cases the world contains an abstract entity, the abstract state of affairs such that Caesar stubs Brutus (and actually also contains such abstract state of affairs even in cases of correct models, so the world would contain the concrete state of affairs Brutus stubbing Caesar *and also* an abstract one Brutus stubbing Caesar). I take it that this is metaphysically odd and should be avoided if possible. And I take it is a virtue of EPS that it avoids it.[11,12]

This point is an advantageous consequence of the tractarian inspiration of the account. Remember that the "representational relation" of the whole model, and the "standing for relation" of its parts, must be neatly distinguished. *And* the global representation relation reduces to the ensemble plus the standing for relation. The representation relation *consists* in the ensemble having its parts (i) in specific behavior to each other and (ii) standing for other entities. As for model-content, once an ensemble has been projected through the standing for relation, this expresses the content of the model; once the world contains an ensemble and a standing for relation on it, it does not contain *in addition* the content of the model. If the model is true, the world contains in addition other ensemble; but if the model is false, the world does not contain some other entity, contains no other thing *at all*.

The second source is specific of standing-for failure. In our last example, the model is incorrect because the stood for entities do not behave to each other as the corresponding pieces of the model behave in the model, but all respects in the model actually stand for the intended entities (in this case of the same logical category). Yet we saw that a model may fail because the intended standing for relation fails, such as phlogiston, caloric and other similar cases. One may defend that in these cases there is no concrete entity stood for, but that nevertheless there is an abstract one (probably claiming also that it is the constituent of the abstract state of affairs). But, what abstract object can it be? Of course, the concept phlogiston exists, as the meaning of the word 'phlogiston' in the theory or model-description; but to claim that the piece in the model stands-for/determines the concept is odd. Much odder if one makes the concept always the stood for entity even when the (aimed!) concrete entity, e.g. oxygen, does exist. And if the abstract object stood for the phlogiston-

[11] For instance, according to Donato and Falguera (2020), there is the model description or material base, the scientific model itself, which they take to be an intensional and abstract entity (an abstract artifact), the target in reality accounted by the model (if any, for they believe that when the model radically fails in postulating theoretical entities that do not exist then there is no target), *and* also an additional abstract entity determined by the model, the (according to them fictional) system *determined* by the model. All those, including the last abstract entity determined by the model, are necessary in order to explicate the nature and function of representational scientific models. It is that abstract entity that I claim my account can dispense with without loss of explicatory power.

[12] I do not claim that this is a virtue exclusive of EPS, Frigg and Nguyen DEKI account for instance is *compatible* with the same non-abstract interpretation; my claim is rather that EPS is the one that makes such non-abstract interpretation explicitly constitutive of the account.

piece of M is not the concept, what other abstract entity can it be? Again, I take this is metaphysically odd and worth to avoid if possible. And, again, I take it is a virtue of EPS that it avoids it.

7.4 Concluding Remarks

According to EPS account, to build a model consists in picking some "pieces" (material or abstract) and ensembling them together constructing a complex/fact (not merely "an aggregate" but an articulated fact). The pieces/components are taken-as-simple, that is, even if complex themselves its inner complexity does not matter for the modeling effects; and in being ensembled/structured to each other, the pieces acquire a logical category in-use. In modeling, this act of constructing an ensemble must come with two other actions, intentional in nature. First, the intention to use the constructed complex to represent a specific target phenomenon previously individuated (by some detached features). Second, the (in general actual, but maybe for some pieces merely fictional) intention of taking the relevant pieces of the ensemble (its "respects") as standing for pieces in the target (including some attributed to the target but not individuative of it). For succeeding in performing the representation, in actually representing (even wrongly), these two intentions must be logically and materially coherent: there must be logical coherence between the logical multiplicity/combinatoriality given to the pieces of the model and the pieces of the target (which does not necessarily imply identical logical category); and there must be material coherency between the intended target and the intended stood for entities, which (the existing ones) must be parts of the intended target, not of "something disconnected else". All these conditions secure the existence of a representation *of a specific target*. In addition, for this representation to be (approx) adequate/correct: (i) the (not fictionally) intended stood for entities must exist; (ii) they must behave to each other C-as (similarity of behavior determined by the context) its originals in the model; and (iii) the modeling purposes are achieved in virtue of such communality of behavior.

Though I could not elaborate it here in detail, it is my claim that this account fares better with respect to explicating performance and adequacy than competing alternatives (inherentisms, functionalisms, representation-as-isms, or fictionalisms). I also claim that EPS permits to dispense unnecessary abstracta. There is a sense in which no account (but radical nominalism, if coherent) can get rid of all abstract entities in modelling, for the construction of the model M uses abstract entities (at least equivalent classes). This is what I call weak abstractionism. But some accounts commit with stronger abstractionism, either accepting possible/fictional systems as abstract entities determined by the model as a whole, or abstract objects as entities stood for parts of the model that aim but "fail" to stand for the aimed concrete entity (caloric, ether, etc.), or both. I take it is an additional virtue of EPS to explicitly dispense abstractionism in this strong sense.

References

Balzer, W., Moulines, C. U., & Sneed, J. D. (1987). *An architectonic for science.* The structuralist program, Reidel, Dordrecht.

Bokulich, A. (2009). Explanatory fictions. In Suárez (Ed.), *Fictions in science. Philosophical essays on modelling and idealization* (pp. 91–109). London/New York: Routledge.

Bueno, O. (1997). Empirical adequacy: A partial structure approach. *Studies in History and Philosophy of Science, 28*, 585–610.

Bueno, O., French, S., & Ladyman, J. (2002). On representing the relationship between the mathematical and the empirical. *Philosophy in Science, 69*, 452–473.

Bueno, O., & French, S. (2011). How theories represent, *Br. J. Philos. Sci. 62*, 857–894.

Burridge, R., & Knopoff, L. (1967). Model and theoretical seismicity. *Bulletin of the Seismological Society of America, 57*, 347–371.

Callender, C., & Cohen, J. (2006). There is no special problem about scientific representation. *Theoria, 55*, 7–25.

Carman, C., Thorndike, A., & Evans, J. (2012). On the pin-and-slot device of the Antikythera mechanism, with a new application to the superior planets. *Journal for the History of Astronomy, 43*(150), 93–116.

Cartwright, N. (2010). Models: Parables v fables. In Frigg & Hunter (Eds.), *Beyond mimesis and convention. Representation in art and science* (pp. 19–32). Berlin/New York: Springer.

Chakravartty, A. (2009). Informational versus functional theories of scientific representation. *Synthese, 172*(2), 197–213.

Contessa, G. (2007). Scientific representation, interpretation, and surrogative reasoning. *Philosophy in Science, 74*, 48–68.

Contessa, G. (2011). Scientific models and representation. In French & Saatsi (Eds.), *The Continuum companion to the philosophy of science* (pp. 120–137). London: Continuum Press.

Da Costa, N., & French, S. (2003). *Science and partial truth: A unitary approach to models and scientific reasoning.* Oxford: Oxford University Press.

Donato, J., & Falguera, J. L. (2020). *The nature of scientific models: Abstract artifacts that determine fictional systems*.

Ducheyne, S. (2008). Towards an ontology of scientific models. *Metaphysica, 9*, 119–127.

Ducheyne, S. (2012). Scientific representations as limiting cases. *Erkenntnis, 76*, 73–89.

Elgin, C. Z. (2010). Telling instances. In Frigg & Hunter (Eds.), *Beyond mimesis and nominalism: Representation in art and science* (pp. 1–17). Berlin/New York: Springer.

French, S. (2014). *The Structure of the World. Metaphysics and Representation.* Oxford: Oxford University Press.

Frigg, R. (2003). *Re-presenting scientific representation*, Ph.D. thesis. London: London School of Economics and Political Science.

Frigg, R. (2010). Models and fiction. *Synthese, 172*(2), 251–268.

Frigg, R., & Nguyen, J. (2016). The fiction view of models reloaded. *The Monist, 99*, 225–242.

Frigg, R., & Nguyen, J. (2017a). Models and representation. In Magnani & Bertolotti (Eds.), *Model-based science* (pp. 49–102). Dordrecht/New York: Springer.

Frigg, R., & Nguyen, J. (2017b). The turn of the valve: Representing with material models. *European Journal for Philosphy Sciences*. https://doi.org/10.1007/s13194-017-0182-4.

Gelfert, A. (2017). The ontology of models. In Magnani & Bertolotti (Eds.), *Model-based science* (pp. 5–24). Dordrecht: Springer.

Giere, R. (1988). *Explaining science: A cognitive approach*. Chicago: University of Chicago Press.

Giere, R. (2004). How models are used to represent reality. *Philosophy of Science, 71*(Supplement), S742–S752.

Giere, R. (2010). An agent-based conception of models and scientific representation. *Synthese, 172*, 269–281.

Godfrey-Smith, P. (2006a). Models and fictions in science. *Philosophical Studies, 143*, 101–116.

Godfrey-Smith, P. (2006b). The strategy of model-based science. *Biology and Philosophy, 21*, 725–740.
Hartmann, S. (1995). Models as a tool for theory construction: Some strategies of preliminary physics. In Herfe, Krajevski, Niiniluoto, & Wojcicki (Eds.), *Theories and models in scientific processes* (Vol. 44, pp. 49–67). Amsterdam/Atlanta: Rodopi.
Hughes, R. I. G. (1997). Models and representation. *Philosophy in Science, 64*, S325–S336.
Hughes, R. I. G. (2010). *The theoretical practises of physics: Philosophical essays*. Oxford: Oxford University Press.
Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese 169*, 615–626.
Jones, A. (2017). *A portable cosmos, revealing the Antikythera mechanism, scientific wonder of the ancient world*. New York: Oxford University Press.
Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science, 42*, 262–271.
Levy, A. (2015). Modelling without models. *Philosophical Studies, 152*, 781–798.
Morgan, M., & Morrison, M. (1999). *Models as mediators. Perspectives on natural and social science*. Cambridge: Cambridge University Press.
Morrison, M. (2008). Models as representational structures. In Hartmann, Hoefer, & Bovens (Eds.), *Nancy Cartwright's philosophy of science* (pp. 67–90). New York: Routledge.
Morrison, M. (2015). *Reconstructing reality*. Oxford: Oxford University Press.
Mundy, B. (1986). On the general theory of meaningful representation. *Synthese, 67*, 391–437.
Peschard, I. (2011). Making sense of modeling: Beyond representation. *European Journal for Philosophy of Science, 1*, 335–352.
Sneed. (1971). *The logical structure of mathematical physics*. Dordrecht: Reidel.
Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of Science, 71*(Supplement), S767–S779.
Suárez, M. (2010). Scientific representation. *Philosophy Compass, 5*, 91–101.
Suárez, M. (2015). Deflationary representation, inference, and practice. *Studies in History and Philosophy of Science, 49*, 36–47.
Suppes, P. (1970). *Set-theoretical structures in science*. Stanford: Stanford University.
Suppes, P. (1974). *The axiomatic method in the empirical sciences*. In Henkin (Ed.), Proceedings of the Tarski symposium. American Mathematical Society, Providence, pp. 465–479.
Suppes, P. (1989). Representation theory and the analysis of structure. *Philosophia Naturalis, 25*, 254–268.
Swoyer, C. (1991). Structural representation and surrogative reasoning. *Synthese, 87*, 449–508.
Teller, P. (2001). Twilight of the perfect model model. *Erkenntnis, 55*, 393–415.
Thomasson, A. L. (1999). *Fiction and metaphysics*. Cambridge: Cambridge University Press.
Thomson-Jones, M. (2011). Structuralism about scientific representation. In Bokulich & Bokulich (Eds.), *Scientific structuralism* (pp. 119–141). Dordrecht: Springer.
Tobin, J. (1970). Money and income: Post Hoc Ergo Propter Hoc? *The Quarterly Journal of Economics, 84*(2), 301–317.
Toon, A. (2012). *Models as make-believe: Imagination, fiction and scientific representation*. Basingstoke: Palgrave Macmillan.
van Fraassen, B. C. (1980). *The scientific image*. Oxford: Oxford University Press.
van Fraassen, B. C. (2008). *Scientific representation: Paradoxes of perspective*. Oxford: Oxford University Press.
Weisberg, M. (2013). *Simulation and similarity: Using models to understand the world*. Oxford: Oxford University Press.
Winsberg, E. (2010). *Science in the age of computer simulation*. Chicago: Chicago University Press.
Wittgenstein, L. (2019). *Trctaus Logico Philosophicus*. London: Routledge (1961).

Chapter 8
The Nature of Scientific Models: Abstract Artifacts That Determine Fictional Systems

Xavier de Donato-Rodríguez and José L. Falguera

Abstract Cartwright (*How the laws of physics lie*. Oxford University Press, Oxford, 1983) said: "A model is a work of fiction". Since then a good deal of philosophical literature has advocated that scientific models are fictional objects. In this contribution we will try to show that Cartwright's *dictum* is correct, but that scientific models are not fictional objects. We distinguish between models as abstract objects and the fictional systems that they determine, where the latter can be in partial correspondence with parcels of the world. We also contrast our view with other recent approaches for scientific models, among them Contessa's dualist view as well as Frigg's, Toon's and Levy's accounts based on Walton's Pretence Theory.

Keywords Scientific model · Fictional system · Abstract object · Abstract artifact

8.1 Introduction

In recent decades there has been a good deal of literature on scientific models (henceforth, SMs) in philosophy of science. This literature has led to a number of prominent distinctions (not always with disjoint extensions). In this sense, various kinds of SMs have been distinguished: *scale models, mathematical models*,

This contribution has received financial support from FEDER/Spanish Ministry of Economy and Competitiveness under the project FFI2013-41415-P, and from FEDER/Spanish Ministry of Science, Innovation and Universities–State Research Agency under the project FFI2017-82534-P. This article also contributes to the project PICT-2014-1741 with financial support from ANPCyT, Argentina.

The order of the authors has no relevance. de Donato-Rodríguez, Xavier and Falguera, José L. have been contributed equally to the elaboration of this chapter.

X. de Donato-Rodríguez · J. L. Falguera (✉)
University of Santiago de Compostela, Santiago, Spain
e-mail: xavier.dedonato@usc.es; joseluis.falguera@usc.es

J. L. Falguera, C. Martínez-Vidal (eds.), *Abstract Objects*, Synthese Library 422,
https://doi.org/10.1007/978-3-030-38242-1_8

analogical models, *theoretic models*, *phenomenological models*, *models of data*, *model organisms*, etc. In any case, what all these different categories of SMs share is that their instances or exemplars are related to the representation of something else: a (set of) possible parcel(s) of the concrete world. Recently, a number of philosophers of science, attempting to understand the representational nature of SMs, have advocated that they are fictional. In saying this, there is sometimes a certain ambiguity revealed in the fact that some of these authors conceive SMs as *works about fictional systems*, whereas other authors tend to understand them as *fictional systems* (see, for example: Godfrey-Smith 2006, 2009; Baberousse and Ludwig 2009; Suárez 2009; Contessa 2010; Frigg 2010a, b; Toon 2012; Levy 2012, 2015; Frigg and Nguyen 2016). As we shall see in the following sections, it seems to us that there is a mistake behind the consideration that an SM is a fiction if we understand this as saying that an SM is a fictional system. In this article, we wish to argue that (i) SMs are not fictional entities and that, notwithstanding, (ii) each SM determines a fictional system. Nevertheless, we advocate that SMs are, in some sense to be specified, analogous to fictional literature and that they should be understood as intensional objects, and, accordingly, as *abstract objects*. Furthermore, we shall argue that the fictional systems determined by SMs should also be considered as abstract objects. More particularly, we contend that SMs and the fictional systems determined by SMs must be considered as abstract *artifacts* owing to their being temporal entities that depend on the formulations and practices of scientists. Finally, we shall endeavour to show that the consideration of SMs (and the associated fictional systems) as abstract objects is appropriate in order to account for the semantic analysis of different kinds of statements about SMs.

8.2 Cartwright's *dictum* and Fictional Entities

Cartwright (1983, 153) claimed that "a model is a work of fiction".[1] The motivation for this claim is because an SM usually incorporates conditions that do not properly correspond to the parcel(s) of the world to which the SM is intended to be applied (as the conditions are highly idealized or may even be incorrect), so that the SM cannot be assessed as providing a veridical (or completely correct) representation of the said parcel(s).[2]

Regarding fictional entities, two main approaches should be distinguished. According to the fictionalist position, the sentences of a fictional discourse–sentences containing expressions about fictional entities (individuals, properties or relations), as with the sentences expressing a novel–do not aim at meaning what is

[1] In a later work, Cartwright (1999, ch. 2) invites us to analyse SMs as fables.

[2] See Godfrey-Smith (2009) for a study of the analogy between fictional systems and fictional characters and for a revision of recent contributions to the analysis of SMs.

literally true, but what is merely pretended to be true. The ontological part of this thesis is that the alleged fictional entities do not exist at all.[3]

The opposite view, let us call it 'fictional realism', is a realist one, which is committed to the existence of fictional entities. Some theorists who advocate for abstract objects contend that fictional entities (e.g. fictional characters) are abstract objects and that abstract objects are really existing entities, though, of course, they have no spatiotemporal existence (at least not spatial existence, possibly just temporal).[4] These theorists are committed to the existence of abstract objects; and, in addition, the fictions themselves–as fictional characters–are abstract objects. In any case, abstract object theorists do not need to affirm that all abstract objects exist in a Platonist eternal world having a kind of privileged existence over and above physical and mental entities, and being independent of any human mind. Abstract object theorists may, like Thomasson (1999), contend that (at least some) abstract objects are constituted objects, depending on human minds which introduce them for different purposes. Thomasson speaks of *abstract artifacts*. And, following Thomasson, we maintain that SMs, and also the fictional systems determined by them, are two sub-kinds of abstract artifacts.

In recent years, it has been usual to find considerations in philosophy of science literature in the sense that sometimes scientific discourse, and other kinds of scientific representations, are, in the first instance, about fictional entities. Among these considerations fictionalist positions are more popular than those committed to the existence of abstract objects. Fictionalist positions were introduced in this kind of literature through authors like Vaihinger (1911) and Duhem (1914), and continued in the work of more recent authors, like van Fraassen (1980), Fine (1993), Baberousse and Ludwig (2009), or Suárez (2009), to quote but a few. Let us focus in the case which is the aim of this article: SMs.

Among the recent philosophers who, assuming Cartwright's *dictum*, have explored the nature of SMs, Contessa (2010), Frigg (2010a, b), Toon (2012), Levy (2012, 2015), and Frigg and Nguyen (2016) have linked the issue to the problems of how SMs represent and how we learn through using them. They are representatives of each of the two main positions on the commitment with fictional entities provided by SMs: we can say that Frigg (with Nguyen), Toon and Levy

[3] Walton (1990) and Everett (2013), among others, provide some important fictionalist accounts on fictional discourse, such as that given by literary contexts. Besides fictionalist approaches on fictional characters, there are fictionalist views on putative abstract entities from mathematics, semantics and so on (e.g., numbers, concepts, propositions, etc.) that reject the existence of such entities. In the case of mathematical discourse, fictionalism has found contenders such as Field (1980), Balaguer (1998), Melia (2000), or Yablo (2001).

[4] Obviously, this requires posing the problem of the semantic evaluation of sentences that contain expressions about such fictional entities. Let us postpone this question for something later; in Sect. 8.3 and 8.5 of this paper.

assume a fictionalist approach regarding the fictional entities provided by SMs,[5] while Contessa adopts a realist one.

For Frigg, Toon and Levy, SMs *are* or *have to do with* fictional or imaginary objects according to a fictionalist approach (and, hence, non-existing entities); they argue in favour of a view based on the Pretense Theory (as applied by Walton (1990) to fictional discourse; e.g., novels). They maintain that SMs (considered as material supports) prescribe imaginings in users of SMs, imaginings that are characteristic of mental processes of make-believe. They refuse to commit to putative fiction entities as existent objects (out of the head).

Frigg (2010a, b) distinguishes several elements which can be considered models in scientific modeling. One of these is the model-system.[6] For Frigg a model-system *is* an imagining obtained by subjects–humans–thanks to a material support (a physical system, a theoretical description or an equation). In this sense, what he calls model-system is a fictional object, a non-existent entity. Toon (2012) and Levy (2012, 2015) prefer to rule out any intermediary between the material support (a physical system or a theoretical description or an equation) and the target system (if the latter exists). Their view is called a "direct" account. Hence, for Toon and Levy, the material support in scientific modeling is the SM, and this material support *has-to-do-with* an imagining for a subject. Toon and Levy discard to speak of model-systems in order to avoid any temptation to consider them as existent. In any case, for all three, modeling through a material support supposes prescribing certain imaginings, which deal with an objective system (when this exists), and imaginings of this kind are fictional in the sense that they are not existing objects.[7]

Contessa (2010) follows a different line. Like Frigg (2010a, b), he speaks of an SM as a putative fictional object: the model-system.[8] Both Frigg and Contessa, in their respective proposals, accept that SMs are (at least) imaginary objects–i.e., that they are analogous to the objects we find in fictional literature. At the same time, they assume that an SM is used to represent a (putative) part of the world (namely, the target system). So that, adopting such an approach, they understand the works of fiction (at least) as fictional (or imaginary) objects. Nonetheless, Contessa proposes

[5]In fact, as far as abstract objects are concerned, Frigg (2010a, b) remains neutral about their existence, though he applies Walton's fictionalist strategy to understand how SMs work; the same can be said of Frigg and Nguyen (2016).

[6]In fact, Frigg (2010a, b) presents what he calls a model-system as a fictional object (or imaginary object). And although a model-system is, in some sense, an SM, he finally adopts an imaginative answer regarding what a scientific model is as he defends the idea that every component, and every combination of components, which he distinguishes in a general picture of scientific modeling (excluded the target-system), can be called *the model*.

[7]Levy (2012) formerly endorsed a direct view too (according to him, there are no intermediate model systems), and he also articulated his view on how scientific modelling works by appealing to Walton's make-believe theory. In a subsequent paper, Levy (2015) defends a fictionalist position as well, while attempting to improve it by appealing to Yablo's notion of partial truth. Our arguments against Levy's position do not differ greatly from those we have against Toon's view.

[8]Contessa speaks of an SM as an imaginary object solely for what he calls 'fictional models'. See Sect. 8.3 below.

a dual ontological nature–*abstract*[9] and *possible concrete*–for SMs.[10] For Contessa, both SMs and fictional characters are imaginary objects and must be considered as abstract objects in order to account for the semantic assessment of sentences about entities of this kind (as we shall see below). In this sense, the SM, as an abstract object, stands for a possible concrete object (a possible target). When the target exists, it is merely an actual concrete object.

Finally, in a recent contribution Frigg and Nguyen (2016) once again maintain that SMs–i.e., what they call model systems–are imagined systems, and develop a new approach called 'DEKI', which attempts to improve on Frigg's own indirect account. According to this new version, an SM represents (denotes) its target (when exists) in such a way that, by exemplifying certain properties, it helps to impute a list of given properties to the target. The SM comes with a *key* to relate the properties exemplified by the model to the properties attributed to the target. For this reason, this account is called 'DEKI' (D for *denotation*, E for *exemplification*, K for *keying up*, I for *imputation*). Although Frigg and Nguyen (2016) have done a great deal trying to improve the indirect version defended in Frigg (2010a, b), we believe that the problems we find in his previous version can be also found in this new approach. Frigg and Nguyen (2016), who once again appeal to Walton's make-believe theory to explain the role imagination plays in scientific representing, maintain that their view is neutral enough to remain outside the scope of metaphysical disputes. In fact, they contend that make-believe procedures can be interpreted as referring to a Meinongian fictional entity, as a way of construing an abstract artifact, or simply as a way of inducing mental contents in the minds of those who are using the model (Frigg and Nguyen 2016, 237). Though this interpretation does not seem us to be compatible with Walton's own approach, which is clearly antirealist as far as fictional characters are concerned,[11] it is certainly coherent to think this way; i.e., one could use Walton's make-believe theory in order to account for scientific modeling and, at the same time, leave the question about the intrinsic nature of SMs open. Nevertheless, the authors are still fictionalists when they say that, typically, in cases in which no material (or physical) models are concerned, imagined objects take the place of the vehicle of representing-as (Frigg and Nguyen, ibid.). So, in these cases model descriptions generate imagined (non-existing) objects that do all the corresponding work.[12] In general, they assume that the imagined (non-

[9]For Contessa, SM and fictional characters are, both, imaginary objects, and these must to be considered as abstract objects in order to account for the semantic assessment of sentences about these kind of entities (as we shall see below). The actual (or real) world is viewed as consisting of both concrete and abstract entities for someone adopting a realist approach regarding abstract entities.

[10]Strictly, Contessa has proposed the dual ontological nature for the kind of SM that he calls "fictional models".

[11]Walton himself would not necessarily be antirealist in relation to numbers or propositions.

[12]We omit Godfrey-Smith's (2006) approach from consideration in this paper, as part of our criticism is also applicable to this perspective. According to his view, SMs are imaginary entities that would be concrete if they existed. On the other hand, and as Morrison (2015, 98) reminds us,

existing) objects are the indispensable intermediaries in scientific modeling. In this sense, despite their apparent ontological neutrality regarding these intermediaries, Frigg and Nguyen favor the interpretation that they are mere imagined objects, as understood from a fictionalist view.

Frigg and Nguyen's proposal also presents some improvements in relation to direct accounts, such as those of Toon and Levy (e.g., the first account provides a better solution for the situation in which there is no target, as in the case of Maxwell's aether[13]), but (as we have said above) there are still several concerns, which–in our opinion–can be equally applied to Frigg's first version. The first issue is that, as we shall go on to say, Frigg and Nguyen's account has to do with how non-existent objects can represent (denote) anything in the world (or even purport to represent/denote). A second issue is related to the fact that, in this view, these non-existent objects are considered to be imaginings. In fact, Frigg and Nguyen's approach, as well as Toon's and Levy's proposals, have to do with Walton's notion of a game of make-believe (and its imaginings) as a fictionalist account for fictional characters, which they attempt to apply in order to explain scientific modelling. If we ask about the ontological status of the imaginings prescribed in the particular game of make-believe which is carried out by the scientist, we face the following issue. Imaginings (like beliefs) are subjective, private entities; so they are not intersubjective. However, SMs are interesting for science only if they are intersubjective entities. After all, the relevant entities for scientific knowledge must be of a public, intersubjective nature. SMs are relevant to science insofar as they have such a condition. It is precisely because they are intersubjective that they are interesting for scientific knowledge (for their scientific understanding).

It is true that advocates of the make-believe approach may say that the intersubjective component, in their proposal about modelling, is given by the material supports as props for imaginings and by the principles of generation that prescribe these imaginings (See Frigg 2010a, 260; 2010b, 264). It must be clear that, in scientific modelling, material supports (physical objects, theoretical descriptions, equations) are not enough by themselves to obtain an intersubjective understanding of SMs; in scientific modelling, a material support needs something else–an interpretation–according to which this material support is understood intersubjectively as a particular way of representing its intended target (whether it exist or not). So, within the make-believe approach, only principles of generation are the relevant component for a truly intersubjective understanding of SMs. Nonetheless, it is not clear what principles of generation are. First, it is not always clear which the principles of generation for each prop are–in our case, for each material support of an SM, as they are not necessarily explicit or clearly identified. Second, it seems that

even if we accepted that it could be possible for SMs to be concrete, this would not affect their condition of being idealisations of their targets.

[13]We should recall that, in the direct approach, a model description, equation, or material entity helps to extract imaginings to represent aspects of the world (a real entity which is the purported target of our modelling). *Only with difficulty can this approach account for the case in which purported targets do not exist.*

the ontological status of principles of generation has to be that of abstract entities, as they need to be objective in order to be intersubjective. The only reasonable option for the defender of the make-believe approach is, or so it would seem to us, to accept that principles of generation are propositions and, hence, abstract objects–appropriately conceived, principles of generation cannot be considered sentences or mere dispositions of the subjects. Now the question arises: if we finally accept propositions as abstract objects, then why not accept SMs as being abstract entities. If something other than mere subjective and private entities are needed, and, given that material supports are not enough and this is not what principles of generation are, then one possible option is to conceive SMs as abstract entities. In fact, Frigg and Nguyen (2016) seem to finally consider this option as a coherent, perfectly possible way of conceiving SMs (although they do not take the step of committing to SMs as abstract entities).

But this is not all that can be said. Fortunately, it is not that considering SMs as abstract entities is a *possible* option; we contend that it is also the *best* one. There are, in effect, reasons that compel us to consider SMs as a particular kind of abstract entity. If we keep in mind a typical work of fiction, for example a novel, it seems natural to think that it is a complex content–i.e., an intensional entity–expressed by its copies, and (as is habitual) that the same content can be expressed by different copies (different material formulations). For a similar reason, and *pace* Toon and Levy, we reject the identification of an SM with the material entity used to express it. In any case, an intensional entity is not a fictional entity. A novel, thus considered, is a cultural abstract product–a set of propositions in this case–which is expressed by each of its copies. Another thing is to say that some intensional objects determine fictional entities. For example, a simple intensional entity, such as the concept of 'hobbit', as established in certain stories by Tolkien, determines a fictional kind; and the complex intensional object which is the novel *The Maltese Falcon*, by Dashiell Hammett, determines a complex fictional situation with different fictional characters, places, facts, etc.[14]

If works of fiction are intensional objects, as we defend, then they are abstract objects. It is usual to consider intensional objects, such as concepts or propositions, as abstract objects; however, one could reject considering SMs as abstract objects arguing that an object, in order to be abstract, must be timeless or independent of human mind. And, so continues the argument, works of fiction, as cultural products, have an origin in and a dependence on the mind of their creator. The answer to this challenge is that it is not necessary to think of this kind of abstract objects as timeless and completely independent of the human mind: it is possible to accept that at least some abstract objects, such as those which concern us here, have a beginning

[14]Here we speak of the relation of determination in a semantic sense, in order to establish what an intensional object is directly about. So, in the case of a fictional content it is directly about a specific fictional object–i.e., a fictional character, a fictional place, a fictional kind, a fictional situation, etc.

(and perhaps an end) and a creator (or several creators). In this sense, they can be conceived as *abstract artifacts* (to use the notion of Thomasson 1999).

Our point of view is that an SM–by virtue of being construed by scientists through a process of abstraction and idealisation–can be considered as a work of fiction, like a novel or a tale. Therefore, we agree with Cartwright's *dictum*, but we understand this in the sense that an SM is an intensional human construction determining an imaginary system, and that it is this imaginary system which is fictional. Nevertheless, it is, of course, a very peculiar work of fiction, as it has the function of purporting some understanding and predictions on some parcel of the real concrete world, on some target system(s).[15]

The last comment leads us to back to the relationship between the imaginary system of an SM and its target system(s),[16] when it exists (they exist). Even if we distinguish an SM as a work of fiction and its imaginary system, given an SM's function of providing understanding and allowing us to make predictions about the target system(s), it seems evident that the imaginary system must have something to do with the target system(s). As we have already said, Frigg and Contessa, and with them others who conceive SMs as imaginary or fictional objects, consider that fictional entities of this kind *represent* parcels of the concrete world. It may seem strange to think that a fictional entity can represent something else out in the concrete world. By using a fictional discourse, we can indeed convey information about the actual concrete world. In novels, we may want to make use of the fictional discourse to provide some information that may correspond, at least partially, to the actual concrete world, as is the case with historical novels and other works of fiction that mix up historical and imaginary situations. But what is really done in these cases, or so it seems to us, is that we are attributing to fictional characters certain properties that correspond to an actual concrete object. This is

[15]In this sense, we should distinguish the case in which a novelist conceives some fictional scientist formulating a new SM–or postulating a new theoretical entity–(where the scientist and the SM construed by him–or the theoretical entity in question–are fictional in the sense that they exist thanks to the novel and its creation by the novelist) from the case in which an actual scientist construes a new SM–or postulates a new theoretical entity–to account for a certain (set of) phenomenon(a). The SM in the novel–let us suppose that it is even conveniently detailed–is not truly a SM, i.e. conceived for scientific purposes, but *just because of the plot of the novel*. Nevertheless, it will be easy to check that our approach can obviously account for these fictional cases as abstract artifacts, because the conceived SM in a novel is determined by that novel in a similar way to a character in a work of fiction (for instance, the Sherlock-Holmes-character of Doyle's stories or the Hobbit-kind in Tolkien's novels). Only in the case where a SM serves for actual scientific purposes and it has been construed by an actual scientist or group of scientists, we are facing an intensional entity that conveniently tries to satisfy scientific standards and that is construed with the purpose of understanding the real concrete world. This proposal is also accounting for cases in which scientists don't believe that the SM stands for a real concrete system (or systems), as for example with the Bohr's atom model nowadays. (The authors of this article want to thank an anonymous referee for his/her helpful comments at this point that lead us to clarify this issue.)

[16]We assume that an SM can have more than an intended target system; for example, Bohr's atomic model has many atoms (especially of hydrogen) as its target systems.

more evident in the case of the relation between the imaginary object determined by an SM and the target system(s) of that SM, as this relation seems to be stronger than the relation provided by any possible metaphorical interpretation. In the case of SMs, some of the aspects of the imaginary object correspond, in fact, to the target system(s) (without any necessary mediation of a metaphorical interpretation to establish them); we can call this “positive correspondence”. It is also clear that certain other aspects of the imaginary object will, in fact, not correspond to the target system(s); let us call this other case “negative correspondence”.

Owing to the positive and negative correspondence, the relation of representation can be called a relation of *partial correspondence*. That is, in saying that the imaginary system determined by an SM *represents* the target system(s) for that SM, we mean that it is a kind of representation that it is not intended to be assessed as veridical. The epistemic value of an SM depends on the implications that positive and negative correspondences (of the imaginary system determined by it) have for the understanding and the predictions about the target system(s). Morrison (2015) has recently criticized Frigg’s fictionalist account by saying that it generally lacks applicability and offers no adequate way to answer the question about the nature of the different types of SMs and how we can learn from them. Her reasons have essentially to do with the fact that fictions, understood in the fictionalist way, work in a very different way than idealizations and abstractions, the real processes that lie behind model construction. Indeed, our account is an alternative way of integrating Cartwright’s idea that SMs are works of fiction and Morrison’s claim that they are the product of a series of idealisations and abstractions. What we contend is that to consider SMs in this way does not imply saying that SMs are imaginings. Of course, every considerable account of SMs also has to explain how, despite their being idealised, we gain knowledge from them about the intended targets. Frigg tries to explain it in his own terms (above all in his 2016 paper with Nguyen). We attempt to improve that elucidation by appealing to a framework that is committed to abstract entities.

Thus far, we have assumed that SMs are works of fiction. More specifically, we have considered them as intensional entities and, therefore, as abstract objects. We have also distinguished between an SM and the imaginary system which is determined by that SM. In addition, we have tried to cast some light on the relationship between the two. Now, before we develop our own approach a bit more, let us consider Contessa’s account in more detail with a view to clarifying certain issues that will be important for the analysis.

8.3 For SMs as Abstract Objects, but Without Dual Nature

According to Contessa (2010), it is possible to distinguish three categories of SMs on the basis of their ontological status: material models, mathematical models and what he calls ‘fictional models’. Paradigmatic examples of each of them (in Contessa’s account) are: a model of DNA that stands on a shelf in a school lab for the first; an equation of the form $dP/dt = rP(1 - P/k)$ for the second; and the ideal pendulum or the Rutherford model of the atom for the third. It is the third kind which he goes on to take into consideration, since "material models do not seem to pose any ontological questions that are not already posed by other material objects" and he "delegate[s] the task of investigating the nature of mathematical models to the philosopher of mathematics" (given its mathematical character) (Contessa 2010, 217).

Contessa calls the third class "the kind of fictional models", because he claims that they are (or are analogous to) fictional entities; in fact, he establishes that they belong to the same ontological genus as fictional entities; namely, that of the imaginary objects. In this sense, he claims that "talking about models and talking about fictional entities are analogous linguistic practices that concern similar kinds of objects but are played for different purposes" (p. 219). Besides, after rejecting the idea that an SM of the third kind might appropriately be an actual concrete object or merely a possible concrete object (or different actual concrete objects or merely possible concrete objects, in different contexts), Contessa considers it as being an abstract entity (pp. 223 and ff.). This option is reasonable as, although an SM of this kind, such as the Rutherford model of the atom, is not a concrete object (a physical system), it does exist in a certain sense given that some sentences–namely, *external sentences*–are literally true or false. External sentences about SMs, as presented by Contessa, are those which are true or false, depending on how the actual world (constituted by both concrete and abstract objects) is (about the creator of an SM, its origin date, etc.), so that there is evidence in the actual world for the truth or falsity of these sentences. For example, an external sentence such as

(1) 'The Rutherford model of the atom was created by Ernest Rutherford at the turn of 20th century'

is true or false for some existing object; and the best way to establish that the Rutherford model of the atom exists is considering that it is an actual abstract entity.

It has to be clear that we agree with the idea that SMs are abstract objects and that this is appropriated in order to perform the semantic assessment of external sentences, as in Contessa's account; however, the reason why we understand SMs as abstract objects is different: it is because we are assuming that they are intensional entities.

In any case, Contessa argues that the view that the SMs of the third kind are actual abstract objects "does not seem to take them seriously enough" (p. 223). The reason is that certain internal sentences are true in some sense, even though they are not literally true about the actual concrete world. We can say that an internal sentence is a sentence about a system as conceived by a work of fiction (in our case an SM); in other words, it is a sentence which attempts to establish how things are *according to* a work of fiction (in our case an SM). For example, an internal sentence such as

(2) 'In the Rutherford model of the atom, electrons orbit around the nucleus in well-defined orbits'

is true in the sense that, according to the Rutherford model, electrons orbit around the nucleus in well-defined orbits; i.e. it is true as assessed according to that model. Of course, the model is wrong, as nowadays we do not accept that electrons orbit around the nucleus in well-defined orbits (in a real atom).

Given this, Contessa holds a dualist account for the ontological nature of the SM, which he calls ‘fictional models’. According to him, these SMs have a dual nature and this dual nature “combines the advantages of both the abstract object and the possible [concrete] object views” (p. 224). But, when Contessa tries to clarify the dualist account, he says that “a fictional model is an abstract object that stands for one or other of a set of possible concrete systems” (p. 224).

Let us indicate that, from the latter characterization of the dualist account, it is difficult to recognise a description of the dual nature of the so-called ‘fictional models’. One thing is to say that these models are abstract objects *and* possible concrete objects, and another that they are abstract objects which *maintain some kind of relation with* possible concrete objects; i.e. the *standing-for* relation (as Contessa calls it). These are different proposals: the first is about a dual *nature* of fictional models; and according to the second, what is assumed is that fictional models are abstract objects and that it is part of their nature to be in the *standing for* relation with systems of a different kind (i.e., possible concrete objects).

In fact, our approach is very close to Contessa’s in this aspect, but it is not identical with it. Where he speaks of two kinds of entities–an abstract entity and a possible concrete object which the former stands for–providing in this sense a dual nature for fictional models, we also say that there are two entities: the SM as an abstract object (because it is an intensional entity) and the imaginary system determined by it. But the imaginary system *is not part of the nature of* the SM. Besides, we conceive the imaginary system as a kind of fictional object. Moreover, because we adopt a realist approach to fictional objects, assuming that they are also abstract objects, the imaginary (or fictional) system determined by an SM is the system for which internal sentences have to be assessed as being true or false. The differences between Contessa’s approach and ours will be addressed in greater detail in Sect. 8.5.

8.4 Scientific Models as Intensional Entities

As we noted above when saying that SMs are intensional entities, we are conceiving them as having a similar nature as other works of fiction (e.g., novels), although with different functions. SMs, like novels, are cultural products. Both are created by intelligent beings. So, both have a dependence on intelligent agents for their initial existence; i.e., both require creators (every work of fiction requires one or more creators or authors). Both have a dependence on a material support; i.e., both require some material support to be expressed as contents. Nonetheless, neither a novel nor an SM needs a specific material support; they just need *some* material support, because either can continue to persist (as intensional entities, as abstract objects) without any specific material support. In both cases different material supports can express them. In the same way as different copies can express the same novel, different material supports (material structures, linguistic descriptions, mathematical descriptions, graphs, etc.) can express the same SM. As we must not confuse a material support for a novel with the novel itself, we must not confuse the material support for an SM with the model itself (in the same vein, we usually distinguish a syntactical expression from its content, a sentence from the proposition it expresses). Both–novel and model–determine a fiction (a fictional world[17] in the case of a novel, and a fictional system in the case of an SM). In any case, the model is not the fictional system that the former determines.[18] This consideration is going to be central to the account we aim to defend here.

Thus, we propose to consider *an SM as an intensional structured entity* which is expressed by means of some material support and which determines a certain fictional system with a given purpose.[19]

---

[17]Here we use "fictional world" in an informal way, accepting that they can be incoherent or incomplete. According to Zalta (1983, 1988), it would be better to call them "fictional situations", because, in his Abstract Object Theory a world is consistent and complete. Zalta himself (see Zalta 2000) makes a useful comparison between Walton's Pretence Theory and his Axiomatic Abstract Objects Theory and considers the notion of story (in his technical sense) as analogous to the notion of a world of fiction (stories are kinds of situations–incomplete possible worlds–that are authored by concrete entities (the writer(s) or the author(s) of the story)).

[18]These comments are inspired by Thomasson (1999), specifically her views on abstract artifacts and their dependences.

[19]J. Díez (at the end of his contribution to this volume) says: "But, what abstract object can it be? Of course, the concept phlogiston exists, as the meaning of the word in the theory or model-description; but to claim that the piece in the model stands-for/determines the concept is odd." And later he says: "But some accounts commit with stronger abstractionism, either accepting possible/fictional systems as abstract entities determined by the model as a whole (cf. Donato and Falguera, this volume), or abstract objects as entities stood for parts of the model that aim but "fail" to stand for the aimed concrete entity (caloric, ether, etc.), or both." These comments are a little bit confusing. Attending to them, in the following we try to clarify our position.

First, we consider that the complex intensional entity expressed by some (different) material support(s) for a scientific model exists, and that this entity is the scientific model. If Díez accepts that concepts as "PHLOGISTON" exists, it is difficult to understand that he doesn't accept complex intensional entities as those expressed by material supports for scientific models. And if he were assuming those entities, why they are not part of his account of scientific modelling?

Second, other researchers, advocating for the make-believe account for scientific models (see the contribution of Salis-Frigg-Nguyen in this volume), are assuming now the necessity to consider *the content* (or intension) of the model description, but they are more interested in the internal process of a human subject related to his imaginations than in the intersubjective content. Against them, we are not very interested in the subjective imaginations (though we don't reject them), because it entails to recuperate the psychologist approach to epistemic–in fact, to ontoepistemosemantics–issues, superseded at the beginning of the twentieth century by several authors within the analytic philosophy (thanks to Frege, Carnap, and Popper, among others).

Third, we do not say that the piece in the model *stands for* the concept (even accepting that the intensional entity which is expressed by that piece is a complex concept, something which we accept). Obviously, that claim is odd, but it is not our claim. It is odd because the verb "to stand for" seems to require the previous existence of the entity which is stood for a piece of material

We shall dissect the last idea with some remarks. First, let us make explicit why we say that an SM is an intensional entity which is *structured*. By this, we mean that it is possible to establish the different sorts of components of an SM: basic notions (or *notions for base domains*) and the different relevant notions for properties and for relations (or *concepts for properties and relations*).[20] An SM is, hence, not a simple entity; rather, it exhibits a certain complexity of greater or lower degree.

Second, an SM determines a fictional system. Every SM can be eventually rejected, and thus considered to provide an inappropriate manner of accounting for how a certain (set of) parcel(s) or system(s) of the actual concrete world is(are). When an SM inappropriately accounts for a certain (set of) parcel(s) of the actual concrete world, it is clear that it is merely determining a fictional system, a system which is not representing a part of the actual concrete world in order to obtain a good understanding of or good predictions about the latter. But if an SM is suitable for accounting for a certain part of the actual concrete world, we can accept that, at best, the SM determines an idealized system having a partial correspondence–a positive and negative correspondence–to that part of the world. If we accept this, we have also to assume that the system determined by an SM is fictional; the system determined by the SM is not the final actual target of that SM. Furthermore, we consider that any fictional object is also an abstract object (although it should be obvious that, for us, not every abstract object is a fictional entity; it should be remembered that we assume that novels, fables and SMs are abstract objects as well, though not fictional entities).

Third, because an SM determines a fictional system, this determined item is structured. Each fictional system so determined imports the structure of the SM; i.e., it is conceived with the structure of the model: it has an extensional structure with domain(s), properties and relations corresponding to the concepts of the SM. This explains why the system determined by an SM can be metascientifically represented by some formal semantic structure(s) or formal semantic model(s). It seems to us that this is what the semantic views of science capture with the idea of models as set-theoretic structures, and that this is what is behind the connection between SMs and models as set-theoretic structures.[21]

---

support (as *representamen*), but in the case of a scientific model–understood as an intensional entity–there is no such entity prior to the first piece of material support (as *representamen*) for the model which is created by its author(s). For us, *the scientific model* (as an intensional entity expressed by some piece of material support) *determines a fictional system*. This latter claim is not odd; it is the usual way to establish the relationship between a *representamen* (especially when it is a linguistic expression) and its content or intension. Only when a corresponding piece of material support expresses a scientific model with an existing target in our actual world it makes sense to speak of "standing for" in order to say that the piece of material support [that expresses a scientific model (as intensional entity)] stands for the corresponding target.

Fourth, in any case, Díez gives a better characterization of our view in the footnote 12 of his contribution to this volume.

[20] In any case, though we assume that these structures have notions as components, they do not necessarily have to be considered as Fregean propositions (or groups of Fregean propositions).

[21] See for example Suppes (1960), van Fraassen (1980), Balzer et al. (1987).

Fourth, a model, as an intensional entity, determines a (set of) fictional system(s) with a given purpose: to account for a parcel or a set of parcels of the real concrete world. This means that the fictional system can have a (partial) correspondence with a parcel (or a set of parcels) of the factual world. But, as alluded to above, this correspondence does not always obtain, or it may be not adequate enough for the purpose of accounting for, an actual parcel of the factual world.

One of the advantages of our proposal is that it includes what Contessa calls 'material models' and 'mathematical models'. In these cases, the concrete materials and the mathematical equations are material supports expressing intensional structured entities. We are not saying that the different ways of expressing SMs are not relevant. The ways of expressing an SM are useful for extracting knowledge from that model, and our proposal includes the different ways of expressing SMs attending to what all these ways express; i.e. intensional entities, and, owing to this, abstract objects. Solely for the purpose of illustration, we can point out that the DNA model is not this or that concrete material support. It is not even the stick-and-ball original material support used by Watson and Crick in 1953 to represent the double-helical structure for the DNA molecule. The original material support for the DNA model has historical importance, but in order to account for the DNA structure and its components there are many other material supports which express (or could express) the same. The content expressed by these different material supports is the common DNA model. This intensional system determines an idealized or approximated system: a fictional DNA molecule. It is supposed that this fictional system obtains a (partial) correspondence with the real DNA molecular structures. The same is true for the case of the mathematical model for the exponential growth of a population. This model is expressed by an equation, namely $dN/dt = (b_0 - d_0) / N_0$, but the equation by itself says nothing. We need the keys of the equation in order to ascertain which model it expresses. Only with these keys will we know what the equation expresses, which is the model. The model is the content expressed by the equation, given the keys for the equation terms. The content, thus expressed, determines a set of fictional population growths.

We began this section by arguing that, in our view, SMs are intensional entities. We also maintained that SMs are cultural products, created by intelligent beings. In order to exist, they depend on intelligent agents and require some kind of material support to be expressed as contents. As we have already mentioned, this position is thus similar to Thomasson's view, according to which fictional characters are abstract artifacts, which are a particular sub-kind of abstract objects. They are artifacts because they are created by human minds, and are not concrete (like chairs), as they lack spatial location, but abstract. The same would be true of SMs. The dependency that abstract artifacts have on the human minds that create them is, in the case of SMs, usually of a theoretical nature. This means that SMs are objects constituted by certain theoretical conditions.[22]

[22]This is something that could be captured by Zalta's Abstract Objects Theory (ZAOT), in which an abstract object is determined by a group of properties, and these properties *constitute* the object

With this account of SMs we can move on to the question of whether our proposal is better suited to solving the problems generated by the semantics of external, internal, and transfictional sentences.

8.5 External, Internal and Transfictional Sentences

We have already argued that it is not the same to speak about a dual ontological nature of SMs as to establish that an SM is an intensional entity which determines a fictional system. We have also said that it is reasonable to treat SMs in a uniform way, despite the different types of material support used for expressing them. Now it is time to see some other advantages of our approach to SMs. One has to do with the two types of sentences related to SMs that Contessa (2010) distinguishes: external and internal sentences. The other advantage has to do with the Pretence Theory as used by Frigg and Toon to account for statements about SMs.

We assume a similar consideration about the semantic assessment of external sentences such as 'The Rutherford model of the atom was created by Ernest Rutherford at the turn of 20th century', or 'The DNA model was conceived by J. Watson and F. Crick, using experimental data collected by R. Franklin and M. Wilkins', or 'The Thomson atomic model is posterior to the model Rutherford atomic model'. The truth or falsity of each of these sentences depends on facts in our actual world. This actual world is not just the actual concrete world. According to the approach assumed in this contribution, the actual world includes concrete entities as well as abstract entities, and among the latter there are objects which are the products of humans (such as novels, scientific theories, SMs, etc., and also the entities conceived according to these works). Because abstract objects, such as SMs, are part of the actual world, it is possible to assess each external sentence as true or false.

Nevertheless, the internal sentences have a slightly different consideration in our approach than that given by Contessa. Sentences such as

(3) 'Electrons orbit around the nucleus in a well-defined orbit',

or

(4) 'The atom is like negatively charged raisins surrounded by positively charged pudding'

---

in question. Abstract objects are introduced in Zalta's theory by means of a *comprehension axiom scheme* which establishes that, for a given condition (or complex of conditions) φ, there is an abstract object that has (encodes) exactly those properties (see Zalta 1983 and 1988). In de Donato and Falguera (2016), we make use of ZAOT in order to present an account of scientific theories, ideal objects and the referents of theoretical terms as special kinds of abstract objects (or more particularly, making use also of Thomasson's ideas, abstract artifacts).

are assessed as false regarding aspects of the actual concrete world.[23] But they have to be considered true in another sense: each of them is true when relativized to the system determined by a certain SM; i.e. to the fictional system determined by the Rutherford atomic model for (3) and to the fictional system determined by the Thomson atomic model for (4). The main difference with Contessa's point of view arises from the fact that, for us, the respective systems for which (3) and (4) are true are fictional objects (according to a realist view about fictional objects in which they turn out to be abstract entities). Contessa says that a sentence like (2) is true in the sense that it is assessed relatively to the Rutherford model. And we agree with him in regard to this consideration. This is clear given the prefix 'in the Rutherford atomic model ...'. In fact, we should do the same with (3) or (4) in order to make the context in which both are true clear; namely, to place the prefix "In ...", or the equivalent prefix "According to ...", before the respective name of the model for each of the sentences; i.e., (2) for sentence (3), and

(5) 'According to the Thomson atomic model, the atom is like negatively charged raisins surrounded by positively charged pudding'

for sentence (4). However, from our point of view, the prefix followed by the name of a model indicates that (3) and (4) have to be assessed in relation to the fictional systems determined by the corresponding SM.

Let us now proceed to explain a further advantage of our approach. This new advantage arises in comparison with Frigg's account,[24] which consists in an application of the Pretence Theory to SMs. In our opinion, his account has a number of problems that concern the semantics of sentences on models and systems in scientific discourse. One is the consideration of SMs as fictional objects. It is not necessary to repeat our arguments against this proposal. Let us just make it clear that, when Frigg speaks about 'model-systems' as fictional objects, he sometimes seems to be speaking about what we call "the fictional system determined by a model" (but without any ontological commitment to fictional objects, as we shall go on to see). If this were the case, then, at first glance, we would not have many problems with what he says about 'intrafictional statements' and 'metafictional statements'. In the case of a novel, for example, an *intrafictional* statement is one which is made within (any copy of) the novel and is considered by a reader, while he is reading, *imaging* that the sentence's content is the case. Therefore, an intrafictional statement is always true. They "are made within the fiction and [the readers of the novel] are not meant to believe them, nor are [they] to take them as reports of fact" (Frigg 2010a, 261). For a novel, a *metafictional* statement is one which can be true or false *in* the novel, as it is asserted by someone speaking about the world of the novel when he is not reading the novel: "Asserting that something is the case in a work of fiction is tantamount to asserting that it is fictional in that work" (Frigg 2010a, 263) (where "being fictional" is the term used by Frigg to say that an intrafictional statement is true in the fiction).

[23] See §3 for sentences (1) and (2).

[24] Toon (2012) adopts a similar position in relation to fictional and metafictional statements.

So, for example, the statement "Holmes is a detective" is an intrafictional one (and hence true) when it is made reading A Study in Scarlet, but it is a true metafictional statement when someone asserts it informing (correctly) about the fictional world in the novel. Nevertheless, it is possible to assert false metafictional statements such as "Holmes is a journalist", because it is not true *in* Conan Doyle's novels. Frigg's solution for assessing some of these sentences as true (or false) is to consider a prefix such as "in the work/(novel) ...".[25] He considers that both types of sentences–intrafictional and metafictional–can be found used in relation to SMs, arguing in that case for the same kind of considerations. It seems to us that both kinds of statements can basically be considered as Contessa's internal sentences.[26] So, our comments above regarding Contessa's internal sentences should also be useful now, especially if, in the end, Frigg is assuming that an SM is a fictional object with no metaphysical commitment to fictional entities.

In any case, Frigg (2010a) is also distinguishing another type of sentence: 'transfictional statements'. These are statements that involve comparing a concrete object with a fictional object.[27] In Frigg's words: "[these] involve comparing something with a non-existent object" (p. 263); for example, comparing some concrete person with Zapp (Zapp being a character in David Lodge's *Changing Places*). Speaking of fictional objects as non-existent objects involves adopting a fictionalist[28] approach about fictional objects, in the sense that they are not abstract entities. It is because of this that, for Frigg, the semantic assessment of sentences of this type is a problem. Thus, he says: "we cannot compare someone with Zapp if there is no Zapp". The solution that he provides is "to rephrase comparative [transfictional] sentences as comparisons between properties rather than [particular] objects". Applying this solution to the sentence 'my friend is just like Zapp' when it is said by Frigg, he claims: "I am not comparing my friend to a non-existent person. What I am asserting is that both my Peter and Zapp possess certain relevant properties which are similar in relevant respects [...] and that these properties are similar in relevant aspects." (p. 263). A similar analysis would be applied to a sentence such as

(6) 'S. Holmes is more famous than A. Pinkerton'.[29]

It should be noted that a paraphrase proposed by Frigg to the last statement would provide a comparison about the degree of fame that, on one hand, is attributed,

[25] In the case of intrafictional statements, the complete operator would be "it is fictional in the work (novel) that ...".

[26] It should be kept in mind that an intrafictional statement could always be asserted as a true metafictional statement. Or, in other way, "*p*, when uttered as a metafictional claim, is true iff *p* is fictional when uttered as an intrafictional claim" (Frigg 2010a, 263).

[27] They can also compare two fictional objects (from different works).

[28] Remember that in a fictionalist approach, fictions are nothing; they do not have any kind of existence.

[29] Allan Pinkerton (1819–1884) was the real detective who was reputed to have discovered a plot to murder A. Lincoln.

outside of the fiction, to a fictional entity and, on the other, is attributed to an entity existing in the actual concrete world. It is important to keep in mind that the present one is not a comparison between the Holmes' degree of fame *in* Conan Doyle's stories and Pinkerton's actual degree of fame.[30] It is a comparison between an abstract artifact created by Conan Doyle, which is fictional *within* the world of some of his novels and is named "Holmes", and the actual concrete object named "Pinkerton". But in order to account for the truth of (6), it is not possible to dispense with the abstract artifact.[31]

Obviously, Frigg is interested in transfictional sentences applied to SMs to account for the comparisons that sentences of this kind establish between what he calls a model-system and its concrete target. Therefore, transfictional statements about SMs, such as

(7) 'The period of oscillation of the bob in the model is within 10% of the period of the bob in the system',[32]

have to be paraphrased (in order to account for their truth) to avoid the commitment to the existence of fictional entities. Applying the solution to SMs, he argues that "truth conditions for transfictional statements [. . .] come down to truth conditions for comparative statements between properties, which are unproblematic in the current context" (p.263). In any case, it seems to us that this solution is obscure, as there is no general criterion for paraphrasing transfictional statements. In our view, it is better to accept that fictional objects–in the general sense, which includes particulars, properties and relations–are abstract objects, because, by assuming this, we have an appropriate semantics for sentences of this kind.

Toon's solution regarding transfictional statements, which is also formulated in terms of the Pretence Theory, is problematic for a fictionalist on fictions, as he himself recognizes. Firstly, it is not clear how to avoid the ontological commitment with the fictional objects to which imaginary properties are attributed and that are compared with those of concrete objects. A second problem arises when we consider sentences such as "The character of Napoleon in *War and Peace* is less well-known as the historical Napoleon". The reason is that, (a) assuming that the sentence is true and that the name "Napoleon" refers in both occurrences to the same real person, the sentence becomes contradictory; and (b) if the name "Napoleon" in the first occurrence refers to a fictional character, then we face the first problem (cf. Toon 2012, 51–53).

[30] Contrast this with the comparison between Holmes and Pinkerton in terms of their relative cleverness, which is made *in the story*. It is far from obvious how a contender of the Pretence View can manage these cases and explain the difference between them. See about this the discussion in Zalta (2000, § 5).

[31] A wider analysis for the problems of paraphrases to transfictional statements according to Pretence Theory can be found in Zalta (2000). See also Godfrey-Smith (2009) for the consideration of some critical comments on Frigg's analysis of transfictional sentences.

[32] The example is discussed by Toon (2012) and also by García-Carpintero (2010).

Problems with transfictional statements are also analysed by García-Carpintero (2010), who adopts a fictionalist approach to fictional entities. His treatment of the problem posed by intrafictional statements is, nevertheless, different from that established by Frigg. García-Carpintero (2010) advocates a figurative version of fictional entities, according to which a metaphorical interpretation is required for statements including fictional expressions–a metaphorical interpretation whose content (in García-Carpintero's view) could not be exhausted with any paraphrase understood as a literal claim. But, in doing so, he is assuming "that many of these transfictional claims do not explicitly make such comparisons" (p. 161), and that "in uttering them, we are not committing ourselves to their truth" (p. 162). However, it is unclear that statements such as (6) and (7) are not uttered in a way that commits their utterers to their respective truths. Nor is it clear that they need a metaphorical interpretation in order for their contents to be well understood. In a case like (6), it seems to us more natural to accept, on one hand, a fictional object as an abstract artifact and, on the other, an actual concrete person. It is, in some sense, more evident that no metaphorical interpretation is needed for (7), as it is asserted given that some mathematical equations expressing scientific laws can be used and applied to an idealized pendulum and to a concrete pendulum, and to compare the periods of oscillation of both bobs: the idealized as abstract artifact and the actual concrete bob.

We find no consideration of external sentences for SMs in Frigg (2010a, b). Maybe this is an oversight, or perhaps they are not important for him. But it seems to us that they are important, in fact crucial, for the issue of the ontological nature of SMs and it is not clear at all how Frigg could account for them. The semantic assessment of these sentences requires assuming that they are abstract objects, as Contessa (2010) argues (and as we ourselves do in this contribution), because they are literally true or false about the actual (abstract and concrete) world. However, remember that Frigg prefers not to take on any metaphysical commitment about fictional or abstract entities. The reason why Frigg does not take on any metaphysical commitment on fictional nor abstract entities is because he is using Walton's Pretence Theory, and he assumes that "[this] theory is antirealist in that it renounces the postulation of fictional or abstract entities" (Frigg 2010b, 120).[33] Frigg is not explicitly advocating an antirealism of that kind, but neither is he rejecting it.[34] He is simply leaving the issue aside in order to understand SMs. The problem is whether this neutrality is adequate and whether it suffices for accounting for SMs. Our comments above try to cast some doubts on the convenience of this neutrality.[35]

[33]Walton's theory is indeed antirealist regarding the issue of fictional entities, though he does not have the same problems with other alleged abstract objects (such as numbers or propositions). See Walton (1990, 390).

[34]Recall here our note 4 above.

[35]We leave the analysis and criticism of Levy's (2015) new account in terms of Yablo's partial truth for a future paper.

8.6 Conclusions

Our approach shows that, though Cartwright is right in saying that models are works of fiction, SMs themselves cannot be seen as fictional entities. In this contribution, we have made sense of the idea that, though construed by a process of gradual abstraction and idealisation, SMs are abstract entities of a peculiar nature. We have seen that SMs are intensional entities and, as such, abstract objects. The fictional systems that they determine are also to be understood as abstract objects, but should not be confused with the former. Finally, we have made sense of the idea that, owing to their being abstract objects, SMs (and the fictional systems they determine) are not necessarily Platonic (timeless) abstracta, because we have taken the option of considering them as abstract *artifacts* in the line of Thomasson (1999). SMs thus come quite close to typical works of fiction and, in this way, Cartwright's *dictum* certainly acquires a certain predicament, but the line of thought that establishes that, in virtue of this, SMs are merely imagined entities is the wrong one. Clearly, we are in no way rejecting the obvious fact that some imagining is generated in the mind of a subject when she is considering an SM (thanks to some material support). The point is that imaginings are not useful for the understanding of scientific intersubjectivity. Hence, we believe that considering SMs as abstract objects in the sense we have advocated for is a better option.

References

Baberousse, A., & Ludwig, P. (2009). Models as fictions. In M. Suárez (Ed.), *Fictions in science. Philosophical essays in modeling and idealizations* (pp. 56–73). London: Routledge.

Balaguer, M. (1998). *Platonism and anti-Platonism in mathematics*. Oxford: Oxford University Press.

Balzer, W., Moulines, C. U., & Sneed, J. D. (1987). *An architectonic for science: The structuralist program*. Dordrecht: Reidel.

Cartwright, N. (1983). *How the laws of physics lie*. Oxford: Oxford University Press.

Cartwright, N. (1999). *The dappled world. A study of the boundaries of science*. Cambridge: Cambridge University Press.

Contessa, G. (2010). Scientific models and fictional objects. *Synthese, 172*, 215–229, 2010.

de Donato-Rodríguez, X., & Falguera, J. L. (2016). On fictions, theoretical entities, and ideal objects: Applying Zalta's abstract objects theory to scientific theories. In G. Borbone & K. Brzechczyn (Eds.), *Idealization XIV: Models in science* (pp. 13–42). Leiden/Boston: Brill/Rodopi.

Duhem, P. (1914). *La théorie physique son objet et sa structure* (2nd ed.). Paris: Chevalier et Rivière.

Everett, A. (2013). *The nonexistent*. Oxford: Oxford University Press.

Field, H. (1980). *Science without numbers*. Princeton: Princeton University Press.

Fine, A. (1993). Fictionalism. *Midwest Studies in Philosophy, XVIII*, 1–18.

Frigg, R. (2010a). Models and fiction. *Synthese, 172*(2), 251–268.

Frigg, R. (2010b). Fiction in science. In J. Woods (Ed.), *Fictions and models: New essays* (pp. 247–287). Munich: PhilosophiaVerlag.

Frigg, R., & Nguyen, J. (2016). The fiction view of models reloaded. *The Monist, 99*, 225–242.

García-Carpintero, M. (2010). Fictional entities, theoretical models and figurative truth. In R. Frigg & M. C. Hunter (Eds.), *Beyond mimesis and convention. Representation in art and science* (pp. 139–168). Dordrecht: Springer.

Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy, 21*, 725–740.

Godfrey-Smith, P. (2009). Models and fictions in science. *Philosophical Studies, 143*, 101–116.

Levy, A. (2012). Models, fictions, and realism: Two packages. *Philosophy of Science, 79*(5), 738–748.

Levy, A. (2015). Modeling without models. *Philosophical Studies, 172*, 781–798.

Melia, J. (2000). Weaseling away the indispensability argument. *Mind, 109*, 453–479.

Morrison, M. (2015). *Reconstructing reality. Models, mathematics, and simulations*. Oxford: Oxford University Press.

Suárez, M. (2009). Scientific fictions as rules of inference. In M. Suárez (Ed.), *Fictions in science. Philosophical essays on modelling and idealisation* (pp. 158–178). London: Routledge.

Suppes, P. (1960). A comparison of the meaning and uses of models in mathematics and the empirical sciences. *Synthese, 12*, 287–301.

Thomasson, A. L. (1999). *Fiction and metaphysics*. Cambridge: Cambridge University Press.

Toon, A. (2012). *Models as make-believe. Imagination, fiction and scientific representation*. New York: Palgrave/Macmillan.

Vaihinger, H. (1911). *Die Philosophie des Als Ob*. Berlin: von Reuther & Reichard.

van Fraassen, B. C. (1980). *The scientific image*. Oxford: Oxford University Press.

Walton, K. L. (1990). *Mimesis as make-believe. On the foundations of the representational arts*. Cambridge, MA: Harvard University Press.

Yablo, S. (2001). Go figure: A path through fictionalism. In P. French & H. Wettstein (Eds.), *Midwest studies in philosophy volume XXV: Figurative language* (pp. 72–102). Oxford: Blackwell.

Zalta, E. (1983). *Abstract objects*. Dordrecht: Reidel.

Zalta, E. (1988). *Intensional logic and the metaphysics of intentionality*. Cambridge, MA: The MIT Press.

Zalta, E. (2000). The road between pretense theory and abstract object theory. In T. Hofweber & A. Everett (Eds.), *Empty names, fiction, and the puzzles of non-existence* (pp. 117–148). Stanford: CSLI Publications.

Chapter 9
The Scope and Power of Abstraction in Science

Stathis Psillos

Abstract In the present paper, I argue that *at least some* models are abstract entities and hence that *at least some* of the content of scientific theories is abstract. I further argue that abstraction, which is ubiquitous in science, is the vehicle for the discovery and representation of abstract objects. Finally, I argue against taking models to be fictions. In Sect. 9.2, I present in some detail the case of the Carnot Engine as a typical case of a model in science and defend the view that it is best seen as a real but abstract entity. In Sect. 9.3, I distinguish and discuss three modes of abstraction in science, which I call Aristotelian, Newtonian and Duhemian. In Sect. 9.4, I offer reasons to resist fictionalism about models. I discuss Kendall Walton's pretence theory of representation and I argue against extending it to the case of scientific models. I show that the current neo-fictionalist account of models fails and that considering models as real but abstract entities is a better account of models than neo-fictionalism.

Keywords Abstraction · Abstract entities · Fiction · Fictionalism · Walton · Aristotle

9.1 Introduction

In 1944, the novelist E M Forster, author of *A Room with a View* and *A Passage to India*, was invited to broadcast a talk at the BBC about the art of fiction. He started with the question 'what is a novel?', and he suggested that the best way to approach it is by bringing to the attention of the audience about half a dozen

S. Psillos (✉)
Department of History and Philosophy of Science, University of Athens, Athens, Greece
e-mail: psillos@phs.uoa.gr

J. L. Falguera, C. Martínez-Vidal (eds.), *Abstract Objects*, Synthese Library 422,
https://doi.org/10.1007/978-3-030-38242-1_9

of novels—pieces that are widely acknowledged as *novels*—and by investigating whether they have anything in common in virtue of which they are novels. This is an admirable task, but he was quick to stress that "the novel is a literary form so wide in its range that generalisations about it are impossible" (1962, 183). In his view, "there is no such thing as the art of fiction. There is only the particular art that each novelist employs in the execution of his particular book". Still, some general things might be said: a novel "should be about human beings (. . .) it should contain a narrative and some sort of plot, and it should be written from a definite point of view". But as he notes, only two of the cases he briefly discussed meet all the 'rules of novel'. And then you might say that only these two are perfect novels, or you might opt for the view that all these are good novels "and leave it like that" (1962, 186).

I want to say that we are in a similar situation when we ask the question "what is a model"? We can cite some examples of models—things that are widely acknowledged as *models*: the Bohr model of the atom; the linear harmonic oscillator; the double-helix model of the DNA; the various iconic models such as the billiard ball model of gases; the Carnot engine; the toy models of molecular elements; the plate tectonic model; the Lotka-Volterra model of predator-prey interaction. Is there anything that makes them all *models*? Perhaps, we can come up with general description of the form: a model represents some aspects of the target system (be they selected parts of the world or the laws and axioms of a theory). We may think of representation as a two-place relation between the model and the target system or as a three-place relation between the model, the target system and the (intentions of the) modeller. But if we dig deeper, it will be hard to unify all undisputed examples of models under a set of rules or features that all and only them share. To paraphrase Forster, the model is *a vehicle of representation* so wide in its range that generalisations about it are impossible.

I am quite happy with the view that a model is what scientists take it to be; and hence with the view that models are vehicles of representation and "and leave it like that". To be a bit more precise, models are part of the content of scientific theories and in particular part of the content by means of which scientific theories represent worldly phenomena and acquire empirical grounding. The focus of this paper, however, will not be the issue of representation. Rather, it will be the metaphysics of scientific models. I have already argued in my (2011) that models (at least *some* models) are best seen as abstract entities. In the present paper, I will explore and further defend this view. Hence, my main thesis is that *at least some* models are abstract entities and hence that *at least some* of the content of scientific theories is abstract. I further argue that abstraction, which is ubiquitous in science, is the vehicle for discovery and representation of abstract objects. Finally, I argue against taking models to be fictions.

Here is a road-map. In Sect. 9.2, I present in some detail the case of the Carnot Engine as a typical case of a model in science and defend the view that it is best seen as a real but abstract entity. In Sect. 9.3, I distinguish and discuss three modes of abstraction in science, which I call Aristotelian, Newtonian and Duhemian. In Sect. 9.4, I offer reasons to resist fictionalism about models. I discuss Kendall Walton's pretence theory of representation and I argue against extending it to the case of scientific models. I show that the current neo-fictionlalist account of models fails and that considering models as real but abstract entities is a better account of models than neo-fictionalism.

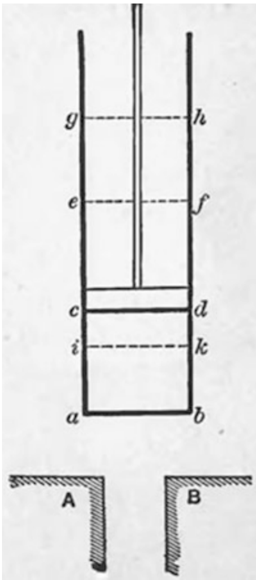
9.2 Models as Abstract Entities: An Example

What are the reasons to take a model to be an abstract entity? To answer this question let us first present and discuss a case of a model as abstract entity. It is by no means unique or atypical, but it is characteristic of the view I want to defend: Carnot's engine.

In standard textbooks, Carnot's engine is presented as a perfectly efficient heat engine. It is introduced by Carnot's theorem: "the reversible heat engine is of the highest conversion efficiency of energy among all other heat engines working between the same hotter heat reservoir and the same colder heat reservoir" (Wang 2011, 3). Sadi Carnot presented it in his (1824) Memoir 'Reflections on the Motive Power of Fire'. He considered two bodies A and B kept at different, but constant, temperatures, T_1 and T_2 respectively, where $T_1 > T_2$. Carnot's cyclic process was to be given for a gas contained in a tank *abcd*, (see Carnot's representation below) whose one side cd is movable with a piston. Carnot's cycle consists in four steps.

(1) The gas is brought in contact with body A, at constant temperature T_1, and is slowly left to expand, at constant temperature T_1, to the position ef.
(2) Body A, then, is removed, and the gas is left to further expand from the position ef to the position gh, until its temperature becomes equal to that of body B, i.e. T_2.
(3) Then, the gas is brought in contact with body B, at temperature T_2, and is compressed from gh to cd, at constant temperature T_2.
(4) Body B is removed, and the gas is compressed from cd to ik, its final temperature being again T_1. Then, the gas is brought in its initial state ab by contact with body A.

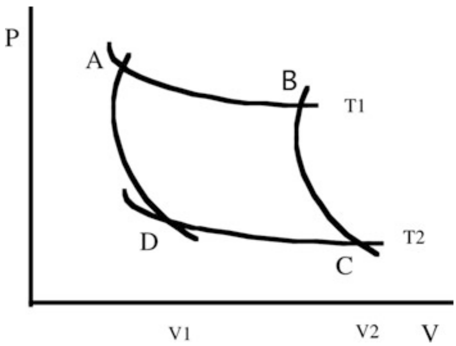
The process could be repeated indefinitely, by repeating the four steps in the same order: (1), (2), (3), (4), ….



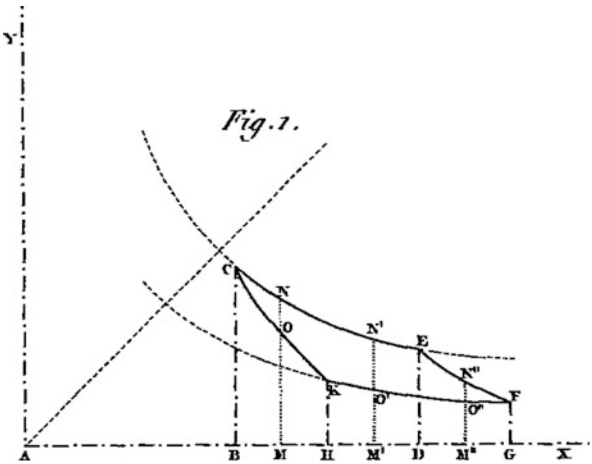
In modern terminology, Carnot's cycle consists in a reversible process, where a gas at an initial state (V_1, T_1) returns to its initial state after having undergone the following successive operations:

(1′) isothermal expansion from V_1 to V_2, (A➔B),
(2′) adiabatic expansion from T_1 to T_2, (B➔C)
(3′) isothermal compression from V_2 to V_1, (C➔D) and finally
(4′) adiabatic compression from T_2 to T_1, (D➔A)

where $T_1 > T_2$.



In fact, this diagrammatic representation was drawn by Benoît Paul Émile Clapeyron in (1834)



Using his cycle, Carnot demonstrated the following:

(1) The *maximum quantity of work* can be produced when and only when a substance undergoes transformations in a Carnot-cycle.
(2) The work produced in a cycle is independent from the substance used and, for a given quantity of heat, depends only on the difference in temperature of the bodies between which the cycle works.

What is the ontic status of a Carnot engine? Here is why the Carnot engine is best understood as an abstract entity. *First*, the Carnot engine (that is the engine that *encodes* the Carnot cycle) is a *type* of heat engine and types are abstract entities. Types, qua abstract entities, have tokens. In the case of the Carnot engine there are no exact tokens. No concrete heat engine can be a perfect exemplification of the Carnot engine, though it might come close to it in efficiency. The fact that it is an uninstantiated type makes it more obvious that it is abstract precisely because it cannot be identified with any concrete heat engine. *Second*, the Carnot cycle does not refer to any concrete heat engine. The reason for this is simple: the physical construction of the reservoirs, the working substance, the method of operation etc. are totally irrelevant to the operation of the Carnot Engine. All that is required is the satisfaction of the description of the Carnot Cycle noted above. Now, we might not want to take the description of the Carnot cycle literally. But *if* we take Carnot's descriptions literally, then there has to be some entity which is thereby described and represented, and this can only be an abstract entity. In fact the different representations noted above are representations of nothing concrete but are *still* representations of something. It is in virtue of this that all these representations are equivalent—this something is an abstract entity: a Carnot engine.

As Carnot himself noted, the working substance of a Carnot engine is indifferent provided we take it to be "susceptible of changes of volume, of successive contractions and dilatations, through the alternation of heat and cold" (1824, 48). In fact, in the operation of the Carnot engine "the necessary condition of the maximum is (. . .) that in the bodies employed to realize the motive power of heat there should not occur any change of temperature which may not be due to a change of volume” (1824, 56–57). The point here is not that this condition is not met in concrete heat engines. Rather, the point is that what goes into a Carnot engine (what Carnot engines are ‘made of’, as it were) are entities that have been stripped of any properties other than those relevant for the reversible operation of the engine. And there is simply no concrete physical heat engine that meets this condition. Using an Aristotelian expression which will be analysed in the next section, in describing the Carnot cycle, Carnot has cut off part of being—the Carnot engine—by abstracting away various attributes of concrete heat engines.

One might wonder: why should we take Carnot’s descriptions literally as purporting to describe something real? To answer it, we should bear in mind that the Carnot engine stands in an important relation to concrete heat engines: it identifies the causally relevant features that a heat engine must have when work is produced; it is a limiting case of a concrete heat engine. In fact, it can be approximated by concrete heat engines, which may satisfy the description to some degree and extent. Recall Carnot’s point noted above: the necessary condition of the maximum work is that in the bodies involved in the engine the changes in temperature should be only due to changes of volume. Carnot added that “This principle should never be lost sight of in the construction of heat-engines; it is its fundamental basis. If it cannot be strictly observed, it should at least be departed from as little as possible” (1824, 56–57). The greater the degree in which the condition is met by actual bodies in concrete heat engines, the better the efficiency of concrete heat engines. Hence, Carnot’s descriptions should be taken literally because they identify causally relevant features of work-producing heat engines.

Still, one might add, wouldn’t we be entitled to treat the Carnot engine as a *fiction*? Now, a lot depends on how we understand fiction (see Sect. 9.3). But if we think of fictions ontologically, it is customary to equate the fictitious with the unreal or non-existent. But then we can see that it is *not* the case that the Carnot-cycle describes nothing. It might not describe any concrete heat-engine, but, as noted already, concrete heat engines may approximate in efficiency the Carnot-engine. In other words, the Carnot-cycle (that is, the associated description) is approximately true of (some) concrete heat-engines. What is the Carnot-cycle (that is, the associated description) strictly true of, then? What is its truth-maker? It cannot be a non-existent entity. Hence, it is something existing but abstract. Even if there are no concrete Carnot–engines, it is true to say that the Carnot engine *is* the most efficient engine. It is also true that if it were not, perpetual motion (of the second kind) would be possible and actual. What makes these truths true is the very Carnot-engine. Could they just be truths-in-fiction (or in–the-story)? But they have worldly consequences: No concrete heat-engine can beat the Carnot-engine in efficiency. How can this be the case if the Carnot-engine is a fiction? The Carnot-cycle sets limits to worldly processes.

Things might get more complicated if we contrast concrete heat engines, located in space and time, with the Carnot engine. The former are hot; the bodies involved in them are concrete bodies with actual volume, temperatures etc. In what sense, the reader might ask, do the bodies of the Carnot engine have temperatures? Is the Carnot engine, to put it crudely, hot? The answer to this worry is: of course it isn't hot. But it would be wrong to conclude from this that the Carnot engine is a fiction. What is true is that the Carnot engine is not in space and time; moreover, abstract entities are not characterised by their predicates in the same way in which concrete entities are. The steam engine of a locomotive is hot in the sense that it instantiates the universal 'hotness'; or if you don't like universals, in the sense that it possesses the trope 'this-hotness'. More generally, concrete entities exemplify their properties. But this is the wrong way to think about abstract entities and their properties. As Ed Zalta (1983) has persuasively argued, capitalising on the work of Ernst Mally (1912), abstract entities *encode* properties in the sense that they are what their properties make them to be. The Carnot engine encodes the Carnot Cycle; it is whatever encodes this cycle. To use Mally's expression, the abstract entity is determined by its properties, whereas a concrete entity instantiates or exemplifies its properties. The Carnot engine does not exemplify having temperature; rather a concrete steam engine of a locomotive exemplifies this property. We may say, however, that the Carnot engine *encodes* the property 'having temperature'. The properties that the Carnot engine encodes are part of its intrinsic—which is its only—nature; they make it what it is.

I have argued in my (2011) that models qua abstract entities are physical abstract entities, that is, they encode physical properties. Their concrete instances, if any, exemplify these properties. Take, for example, the Linear Harmonic Oscillator (LHO), which qua abstract entity, is not to be found in any spatio-temporal location. A LHO is an object which satisfies a certain equation, viz., its motion (along x axis) satisfies the second-order differential eq. $F = m\, d^2x/dt^2 = -kx$. It is characterised by physical predicates, e.g., motion, mass, force. Being an abstract object, the LHO does not move; that is, it does not exemplify motion (or mass, or force etc.). Yet, it encodes these properties in the sense that it is exactly this entity in which these properties are related to each in the exact way described by the foregoing equation.

Note, by the way, that this issue of the distinction between exemplification and encoding is orthogonal to the distinction between models-as-abstract entities and models-as-fictions. A fictionalist about models would do well to respect it. Here is why. A fictionalist about models goes, typically, for a literal understanding of the descriptions associated with a model. Their key point is that since nothing concrete satisfies the description, they refer to nothing. Or they refer to an "imaginary object". Now, it is true that given that the descriptions of models involve idealisations and abstractions, nothing in the world satisfies the associated descriptions. Hence, nothing in the world exemplifies, strictly speaking, the properties talked about in the associated descriptions. But since it is a contingent matter that nothing in the world satisfies the associated descriptions, a fictionalist should certainly allow that it is possible that the descriptions of the model might be satisfied. Whatever possibly satisfies the (highly idealised and abstract) associated descriptions—whatever that is, satisfies the literally understood associated descriptions—cannot possibly exemplify the relevant properties; hence it should be said to encode them. The difference between the fictionalist and the advocate of the models-as-abstract entities is not over whether models encode their properties. Rather, the difference is over a further issue, viz., whether or not the fact that nothing concrete exemplifies these encoded properties should or should not have a bearing on the ontic status of models. The advocate of models-as-fictions says that lack of exemplification implies that models are fictions, whereas the advocate of the models-as-abstract entities denies this.

If all this is in the right direction, Carnot discovered the Carnot-cycle and hence the process of outlining its operation is a process of discovery of an abstract entity: the Carnot-engine. How is this possible? That's exactly where abstraction plays a key role. Abstraction is the means to discover and to characterise abstract objects.

9.3 Three Types of Abstraction

Models as abstract entities rely on abstraction for their existence and representational capacities. In this section, I will present three ways in which abstraction is the vehicle for the discovery of abstract objects. I will call them: Aristotelian, Newtonian and Duhemian.

9.3.1 Aristotelian Abstraction

I call the first kind of abstraction "Aristotelian" because it can be traced back to Aristotle's idea that each science investigates some part of being by isolating the relevant attributes or properties of the elements of this part. So, we can define Aristotelian abstraction as cutting off parts of being and investigating the attributes of these parts. In *Metaphysics*, Aristotle define first philosophy thus:

> There is a science which investigates being as being and the attributes which belong to this in virtue of its own nature. Now this is not the same as any of the so-called special sciences; for none of these others deals generally with being as being. They cut off a part of being (ἀλλὰ μέρος αὐτοῦ τι ἀποτεμόμεναι) and investigate the attributes of this part—this is what the mathematical sciences for instance do (1991).

For Aristotle, metaphysics (*philosophia prima*) studies being qua being and each individual science studies parts of being. My point of interest though is not in the distinction between metaphysics and the sciences. It is, rather, in the very idea of cutting off parts of being; that is of abstracting parts of being and studying their attributes. Abstraction is not just the process of cutting off. It is also the product: the entity left over after the cutting off and its attributes. Hence, abstractions are parts of being: an abstraction is this part of being that is left over when various features of the being are eliminated. Aristotle used the example of the mathematical being, which is the (part of being) which is the product of the elimination of all features other than "the quantitative and continuous" and the consideration of only "the attributes of things qua quantitative and continuous". Here is how he put it:

> as the mathematician investigates abstractions (for in his investigation he eliminates all the sensible qualities, e.g. weight and lightness, hardness and its contrary, and also heat and cold and the other sensible contrarieties, and leaves only the quantitative and continuous, sometimes in one, sometimes in two, sometimes in three dimensions, and the attributes of things qua quantitative and continuous, and does not consider them in any other respect, and examines the relative positions of some and the consequences of these, and the commensurability and incommensurability of others, and the ratios of others; but yet we say there is one and the same science of all these things—geometry) 1991.

If abstractions are part of being, they are real entities. But they are abstract entities since they lack various features that would make them concrete entities. The case of mathematical entities is clear and uncontroversial. On the Aristotelian account, they are abstract entities because they lack all sensible qualities. This does not mean that they lack observable qualities. Rather, the point is that they lack physical qualities—qualities in virtue of which they can engage in a causal give-and-take with other entities.

But there are physical abstract entities too. That is, abstract entities which encode physical properties. When, for instance, Carnot considered the bodies of the Carnot cycle as having only volume and temperature, he 'cut off' part of being: no concrete physical body has only these two attributes. The resulting abstraction is a body qua subject to a Carnot-cycle of operations. There is no concrete body like this. Hence, it is an abstract entity, which however encodes physical properties: those that are relevant to its being subject to a Carnot-cycle. After all, the product of this process of cutting off parts of being, the Carnot-engine, is the very entity that satisfies the Carnot-cycle; and hence an abstract entity itself.

The process of cutting off parts of being has been described by Allan Bäck as selective attention. He characterises selective attention as "focusing on an aspect, typically a general one, and then looking at features belonging to that aspect, while ignoring the remaining ones" (2014, 2). But as noted already, abstraction can be both a process (selective attention) and product (the objects being selectively attended to). In this sense, abstracta are new objects: they might be generated by abstraction or they might be *recognized* by abstraction. The Carnot-engine is the recognition of a new object, which sets limits to the efficiency of concrete heat engines.

As Bäck correctly notes, abstracta "have properties of the same type as those that the original substances have" (2014, 23). Aristotle used the example of the triple: noses, snubness and concavity. Snubness is an abstraction from real concrete noses. Concavity is an abstraction from snubness, which is a property of noses. It can be thought of as form abstracted from matter: "Of things defined, i.e. of essences, some are like snub, and some like concave. And these differ because snub is bound up with matter (for what is snub is a concave nose), while concavity is independent of perceptible matter" (1025b22–25).

To sum up. Aristotelian abstraction is selective attention, which cuts off parts of being by eliminating features of the being, thereby recognizing new objects.

These objects are real and objective though not independent of concrete (and perhaps sensible) particulars in that they could not be there if all particulars were eliminated (hence, they are not Platonic Ideas). These objects encode the same type of properties that concrete objects exemplify, though only some of them—those which permit their study independently of any particular instance (token) of them.

9.3.2 Newtonian Abstraction

Here an extra layer of abstraction is added: laws and relations among abstract entities. The abstract objects (qua cut-off parts of being with specific general features) are subjected to mathematical treatment; they become *magnitudes*. Hence, in Newtonian abstraction an abstract entity is considered insofar as it stands to relations with other entities. Roughly put, Newtonian abstraction establishes propositions about the relations among physically abstract entities; and more specifically about the laws that capture their relations.

Let us illustrate this by an example from Newton. Take Proposition 1, Theorem 1 of Book 1 of the *Principia*:

> The areas which revolving bodies describe by radii drawn to an immovable centre of force do lie in the same immovable planes, and are proportional to the times in which they are described (1729, 40).

Newton, as is well known, offered a purely geometrical proof applied to objects (*revolving bodies*) whose only properties are those relevant to their motion (revolution). The law (or half of the law, to be precise: if there is a centripetal force, then the area law is satisfied) expresses an exact relation among abstract entities: revolving bodies, radii, immovable centre, immovable plane, force.

It is noteworthy that Newton introduces this idea of a generic sense of force, which as he put it, is mathematical as opposed to physical. By this he means that forces are considered as magnitudes, being cut-off from any particular mode of operation ("manner of action") or from any actual physical location. The forces studied by Newton, in Book 1, are physically characterised abstract entities, which obey certain laws (that is, they stand in relations to other abstract entities). He warned his reader that he (Newton) considers

> those forces not physically, but mathematically: wherefore the reader is not to imagine that by those words I anywhere take upon me to define the kind, or the manner of any action, the causes or the physical reason thereof, or that I attribute forces, in a true and physical sense, to certain centres (which are only mathematical points); when at any time I happen to speak of centres as attracting, or as endued with attractive powers (1729, 5-6).

When the nomological propositions are applied to the phenomena it becomes even clearer that they are 'about' abstract entities. Consider the following statement made by Roger Cotes in the Preface to the second edition of Newton's Principia in 1713: Statement X

> Therefore the sun gravitates towards all the planets, and all the planets towards the sun. For the secondary planets, while they accompany the primary, revolve the meanwhile with the primary about the sun. Therefore, by the same argument, the planets of both kinds gravitate towards the sun and the sun towards them (1729, xxv).

Proposition 1 above is about "an immovable Centre of force" while, according to statement X, "the sun gravitates towards all the planets"; hence it is not immovable. But there is no contradiction here. Proposition 1 is about the relations among abstract entities and is proved of them. Statement X is about the actual motion of concrete entities (the sun and the planets). The centre of force (the Sun) is not immovable, since it is actually subjected to the gravitational attraction of the planets. Still, the abstract relation described by proposition 1 is 'close enough' to the concrete relation described by statement X. The abstract relation holds for the concrete world *quam proxime*. This phrase (which occurs 139 times in the *Principia*) literally means 'most nearly to the highest degree possible'; it is probably best translated 'very, very nearly' or, as is more customary, 'very nearly'. Hence, what accounts for the applicability of the law that holds among abstract entities to the worldly phenomena is that the concrete physical system satisfies the law *quam proxime*. In other words, the law is encoded by abstract entities and is satisfied *quam proxime* by concrete ones.

9.3.3 Duhemian Abstraction

Duhemian abstraction offers an answer to the question: How are new objects (qua abstract) regognized/introduced? Duhem's key idea is that they are introduced by abstraction principles.

Here is his example. Warmth is an empirical property of bodies. Bodies can be as warm as others or more or less warm than others. These features, however, though "essential to the concept of *warmth*, do not permit the measurement of the object of this concept—that is, to regard it as a *magnitude*". And yet, the relation of *being as warm as*, which holds between actual physical bodies given in experience, has the properties of being reflexive, symmetric, and transitive. It is, in modern terminology, an equivalence relation. Duhem observes that a more rigorous physical concept (and a corresponding magnitude) can be introduced on the basis of this equivalence relation, viz., temperature. As he (1892, 3) put it:

> We make two equal values of temperature correspond to two points that are as warm as each other. We make two unequal values of temperature correspond to two points that are not equally warm, and in such a manner that the higher value of temperature corresponds to the warmer point.

This move allows, among other things, the transition from a qualitative property to a quantitative one. Even though it makes sense to say that body A is as warm as, or warmer than, body B, it does not make sense to assert that the warmth of the body C is equal to the warmth of body A plus the warmth of body B. Not so for temperature, since this is a magnitude which is additive. In fact, the concept of temperature has excess content over the concept of warmth, since the concept of temperature involves an extra quantitative assumption, according to which to each point of a body can be assigned a definite value of temperature—an assumption that permits additivity.

The concept of temperature, then, is introduced on the basis of an *abstraction principle* over a set of physical bodies and an equivalence relation among them (*being as warm as*). Duhem is clear that this procedure is essentially generalizable. What was said about temperature "could be repeated—at least in its essentials—about all definitions of magnitude that we find at the beginning of any physical theory whatsoever" (1892, 4). Note that this is a far-reaching approach. It shows that Duhem was appealing to what came to be known as abstraction principles in order to introduce new physical concepts and corresponding magnitudes.

This definition of physical magnitudes by abstraction out of equivalence relations among the properties of empirical bodies makes possible the statement of mathematical-quantitative physical laws about the magnitudes thus defined. In particular, it makes possible the symbolic representation of empirical laws, even if these laws do not bear strictly speaking on the same empirical properties as the empirical laws did. For Duhem, abstraction is indispensable because the magnitudes to which the theoretical hypotheses of a physical theory apply must be mathematized, so that these hypotheses state mathematically precise relations (laws) among magnitudes.

The logic of abstraction principles was developed by Russell (1903) and Frege (1884, §§63–67). How do they work? As Russell noted, whenever a relation is symmetric and transitive (an equivalence relation), the relation is not primitive but "analysable into sameness of relation to some other term" (1903, 166). This analysis leads to the specification of a "common property" of the two terms; this is a new third term to which "both have one and the same relation" (1903, 166). His prime example is the definition by abstraction of *cardinal number*. Here is how it works.

1. Number may be regarded as the property of a class. When any class-concept is given there is a certain number of individuals to which this concept is applicable.
2. Two classes have the same number when their members can be put in 1–1 correspondence.[1]
3. When two classes have the same number they are similar.
4. Similarity is a relation that is symmetric and transitive (and reflexive).
5. When the relation of similarity holds between two terms (two classes), those two terms "have a common property and *vice versa*" (1903, 114).

[1] "Two classes have the same number when, and only when, there is a one-one relation whose domain includes the one class, and which is such that the class of correlates of the terms of the one class is identical with the other class" (1903, 113).

6. The number of class X is the same as the number of class Y iff X is similar to Y.[2]

More generally, an abstraction principle has the form:

$$\text{For all x for all y } S(x) = S(y) \text{ if } R(x, y)$$

where S is term forming operator (i.e., a function from elements of the initial set of entities G to the abstract entities of set S) and R(x, y) is equivalence relation over the members of the initial set G.[3]

It follows from the use of Abstraction Principles that abstract objects are not creations of the human mind. They are already 'present' in the right-hand side of the Abstraction Principle and the Abstraction Principle shows how the right-hand side can be *reconceptualised* so that part of its content is seen as being about abstract objects. Hence, the issue of the existence of abstract objects is co-ordinated with the issue of the existence of the entities in the right-hand side of Abstraction Principle. The latter introduces a new concept and specifies the conditions under which objects that fall under this concept are the same.

When talking about the introduction of direction, Frege (1884, 75) noted:

> Thus we replace the symbol // by the more generic symbol =, through removing what is specific in the content of the former and dividing it between a and b. We carve up the content in a way different from the original way, and this yields a new concept.

This Fregean idea of carving up the content is akin to the Aristotelian idea of cutting off part of being. They share in common that abstract objects might well depend for their existence on concrete ones, but they are not reducible to them.

We have already seen Duhem introducing magnitudes by abstraction principles. What Duhem emphasised was that abstraction makes the mathematisation of nature possible. As he put it: "It is abstraction that furnishes the notions of number, line, surface, angle, mass, force, temperature and quantity of heat or electricity. It is

---

[2] In *Grundlagen*, he showed the power of abstraction principles by introducing the concept of direction over the equivalence relation 'being a parallel line to' (1884, 74–78; *§63–67*).

(D=) The direction of the line *a* is the same as the direction of the line *b* if, and only if *a* is parallel to *b*.

or,

(D=) D(a) = D(b) iff a//b.

For Frege, lines are given in (spatial) intuition and yet directions (introduced as above) are abstract entities *not* given in intuition. The concept of direction is discovered by a process of intellectual activity which takes its start from intuition. For Frege abstraction principles explain our capacity to refer to abstract objects. 'The direction of the line *a*' is a singular term; it refers to an object. (D=) enables us to identify this object as the same again (criterion of identity) under a different description, e.g., 'The direction of the line *b*'. We thereby have a criterion of identity: the criterion of identity of the new entities S (e.g., directions) is given in terms of a relation R (which is an equivalence relation) on things of some other kind G (e.g., parallel lines).

[3] For a succinct but informative account of abstraction principles see Horsten and Leitgeb (2009).

abstraction, or philosophical analysis, that separates and makes precise the fundamental properties of these various notions and enunciates axioms and postulates" (1893, 58).

9.4 Models as Fictions

So far, I have focused my attention on giving reasons for thinking that a class of models, an example of which is the Carnot engine, are abstract entities. I have also suggested that abstractions (of which I have discussed three types) are means for discovering and representing abstract objects. But this is not the only view available.

A popular view is that models are fictions. But what exactly is a fiction? And is fictionalism a viable alternative to abstractionism?

9.4.1 Traditional Fictionalism

Generally speaking, fictionalism is the view that some entities whose existence is implied by the truth of a theory are not real, but useful fictions. Hence, the relevant theory is false. On this view, to say that one accepts the proposition that *p as if* it were true is to say that *p* is false but that it is useful to accept whatever *p* asserts as a fiction.

This view was historically introduced by Hans Vaihinger in his *The Philosophy of As If* (1911). Vaihinger noted that what is meant by saying that matter consists of atoms is that matter must be treated *as if* it consisted of atoms. But what does it mean that "matter must be treated as if it consisted of atoms?". As he said: "It can only mean that empirically given matter must be treated as it would be treated *if* it consisted of atoms" (1911, 93). Though it is false that matter has atomic structure, Vaihinger argued that the as-if operator implies a decision to maintain formally the assumption that matter has atomic structure *as a useful fiction*. Hence, we may willingly accept falsehoods or fictions if this is useful for practical purposes or if we thereby avoid conceptual perplexities. We then act *as if* they were true or real.

It should be noted that Vaihinger's fictions are not *just* false assumptions—they are knowingly false and, in their stronger version, impossible to be true. Vaihinger distinguished between fictions and hypotheses—the latter can be true or false and it is an open issue what they are; the former *cannot* be true. Fictions "lead to contradictions" Vaihinger says. The as-if formula, he says, "states that reality as given, the particular is compared with something whose impossibility and unreality is at the same time admitted" (1911, 93). According to fictionalism, then, fictions are not real entities; fictions don't exist.

Could it be that models are fictions in Vaihinger's sense because they rely on abstractions? That would be the wrong way of thinking. We have already seen that abstractions are the products of cutting off part of being. By cutting off part of being, you may end up with something abstract but not unreal! Universals rely on abstraction. A universal is the one over the many. As such, it recurs in different spaces at the same time. It cannot possibly be a concrete entity because no concrete entity is (or can be) at different places at the same time. Besides, the universal abstracts away all but one feature of the various particulars that exemplify it. I am not saying that universal must be admitted. My point is simply that the very fact that they are abstract does not ipso facto make them fictions. It should also be noted that concrete entities might rely on abstractions without thereby being fictions. Maps are concrete entities, but they represent by abstracting away various features of the terrain (including its actual size). Clearly, maps are not fictions. Hence, the very fact that models rely on abstractions does not imply that models are fictions. Are there other reasons to take models to be fictions?

9.4.2 Pretence Theory: Models as Make-Believe

A rather popular view of models is to think of them as make-believe. This view capitalises on Kendall Walton's (1990) account of representation in the representational arts. Let us take a look at this theory.

Walton takes it that the key to understanding representation in art is held by typical children's make-believe games. Think of Alice and Bob intending to play a bear game. Alice and Bob go out in the woods and decide to use the stumps as a prop for bears. They *imagine* that the stumps are bears. And they play along. 'Here is a bear' cries out Alice. 'Be careful; she is dangerous' replies Bob; and so on. This act of imagination is essential for the pretence theory. Walton notes that a make-believe game involves a *mandated imagining*: those who play it have to imagine that the chosen props represent what they are supposed to. This is ensured by a principle of generation, according to which prop X represents Y. A make-believe story involves a fictional world. If within the bear game, Alice sees a bear (that is, if she sees a stump which is imagined to be a bear), then it is true in the fictional world that there is a bear. This is what Walton calls "fictional truth". They are generated by props (cf. 1990, 40). The role of props in pretence theory can hardly be exaggerated: no props, no representation. Walton sums up his theory thus:

> *Representations,* (...), are things possessing the social function of serving as props in games of make-believe, although they also *prompt* imaginings and are sometimes *objects* of them as well. A prop is something which, by virtue of conditional *principles of generation,* mandates imaginings. Propositions whose imaginings are mandated are *fictional,* and the fact that a given proposition is fictional is a *fictional truth. Fictional worlds* are associated with collections of fictional truths; what is fictional is fictional in a given world—the world of a game of make-believe, for example, or that of a representational work of art (1990, 69).

Works of fiction (and representational arts in general) involve propositions which are "true in a fictional world", that is propositions which are imagined (as true) in a given game of make-believe. But Walton is very careful to distinguish fictionality from truth. To say that a proposition is fictional is to say that imagining it is prescribed by a rule of generation in a game of make-believe. Far from ascribing a property of truth to it, a totally distinct property is ascribed, viz., fictionality (cf. 1990, 42).

Does all this imply that there are fictitious entities? A fictitious story is about "*mere* fictions" (1990, 73). But what are they? Walton has no sympathy for either what he calls "voodoo metaphysics", according to which some lesser kind of existence is attributed to fictions, or to the view that fictions are abstract entities (cf. 1990, 385–387). The key reason for objecting to treating fictions as abstract objects is that "we describe fictions as we would ordinary concrete particulars" (1990, 387). This is obvisously too quick: Sherlock Holmes is described as a concrete object but the Carnot engine, as we have already seen, is not. Concrete heat engines have reservoirs made of certain material with well-defined dimensions, to say the least. A Carnot engine does not!

Be that as it may, Walton is fully aware of the key problem faced by the pretence theory, qua a fictionalist theory. To say that something is a fiction is to say that it is non-existent. But then what do we assert when we assert that e.g., King Lear has three daughters? To what do we attribute the property of having-three-daughters? Or, to use our own case, to what do we attribute the property of being-the-most-efficient-heat-engine if the Carnot engine is non-existent? As Walton put it: "Why should we disguise what are really observations about the *nature* of certain objects as denials that there are such?" (1990, 388). But having acknowledged the problem, Walton aims to dissolve it.

His dissolution is long and intricate. But the key issue is that Walton insists that the practice of talking *about* fictional entities does not yield any ontic commitments to there being any of them. Given our immersion to a make-believe game, we mistake the *pretence* of referring to fictitious objects with being ontically committed to them. Hence, there are no objects to which fictional propositions are committed to. But then the issue of their reality, e.g., the issue of their being abstract objects, does not arise (cf. 1990, 390).

How then should statements such as 'Gulliver was captured by the Lilliputians', or 'The Carnot engine is reversible' be understood if they are not about anything? And as Walton admits, is it not simpler to take them as making genuine assertions? His answer is that they should be seen as locating the speaker *within* a fictional world and, in particular, as contributing to it (cf. 1990, 392). Seen this way, what makes a claim such as 'Gulliver was captured by the Lilliputians' true has nothing to do with Gulliver's being, in some sense, real. Rather, it has to do with the fact that for a subject to engage in an act of pretence while participating in a game of make-believe about the novel *Gulliver's Travels* is for this subject to "fictionally speak the truth" (1990, 400). Hence, Walton dispenses with fictional entities and transfers the burden of assertion to the participant of the game: what the participant asserts from within the game (novel) is true.

Can this manoeuvre succeed in avoiding commitment to entities? Walton is adamant that it can succeed only if the paraphrase he offers is adequate. But he (1990, 404) expresses doubts about it. The transition from thinking about the truth-conditions of a perfectly ordinary English statement such as 'Gulliver was captured by the Lilliputians', or 'The Carnot engine is reversible' to thinking about principles of generation in games in which participants "speak the truth" is hardly straightforward. Nor does it eliminate the apparent references to entities such as Gulliver or the Carnot engine. Indeed, it's hard to see how a commitment to entities should be avoided by means of the claim that the participant speaks truly. The content an assertion such as 'Gulliver was captured by the Lilliputians' might have is not captured by saying, in effect, that this proposition is fictive in a certain world. For what makes this assertion be *about* Gulliver (and not someone else) if not that it is made true by Gulliver? To say, in effect, that its truth is implied by the novel is not enough unless the novel is about Gulliver. Walton acknowledges the difficulty but he does not give in. He takes it that "it is our pretendings to assert, our games of make-believe, that are central to our conceptual scheme. It is that, not an ontological commitment to fictional entities, that plays an important role in our structuring of the world" (1990, 404).

Even if we were to grant this for the case of novels, it's hard to see how it could be extended to scientific models. Commitment to the Carnot engine does play an important role in structuring the world, since, among other things, a limit to the actual efficiency of heat engines is set. If this is true, it is not so because it is asserted by whoever adopts Carnot's theory but because it captures a fact about an entity, which albeit abstract, it does resemble to a good extend, real concrete entities.

More generally, it might be doubted that Walton's theory, despite its general interest, is suitable when it comes to scientific models. The attraction of Walton's theory, according to Adam Toon (2012, 74), is in the thought that models "function as props in games of make-believe". Hence, they are not works of fiction in general but in "Walton's sense", meaning that models mandate imagining what the model says. Models, as Toon puts it, "prescribe a web of imagining which the scientist can then go on to explore (2012, 75). What I would doubt is the very idea of imagining. I don't know, for instance, whether Carnot imagined or not the Carnot cycle, but I think this is irrelevant to the function of Carnot engine as a model of a heat engine. When Carnot engaged with the Carnot cycle, he asserted its salient properties and proved that nothing can match it in efficiency. Indeed, he did explore the cycle but not by putting it in an network of imaginings but by drawing conclusions about the motive power of heat. What matters for the functioning of the model is the inferences about concrete systems that makes possible.

9.4.3 Neo-fictionalism

Though Walton did not talk about models, his pretence theory has been recently adopted by philosophers of science as the theoretical basis of a fictionalist account of models. There are at least two approaches here. One is by Toon, who takes it that the model (qua make-believe) represents *directly* the target system; hence it is *about* it. For instance, a LHO (better, the equations of motion and the theoretical descriptions associated with it) represents directly concrete pendula and not an abstract intermediate (an ideal or abstract oscillator). What exactly is fictional about this view? That it is said that the model represents the concrete physical system by mandating imaginings about it, e.g., the bob of the actual pendulum is *imagined* as being a point mass; hence, it is fictional that the bob *is* a point mass.

The other approach is due to Roman Frigg, who has pioneered neo-fictionalism. He takes it that models are imagined concrete things, i.e., fictional concreta. He says: models are understood as

> imagined physical systems, i.e. as hypothetical entities that, as a matter of fact, do not exist spatio-temporally but are nevertheless not purely mathematical or structural in that they would be physical things if they were real (2010a, 253).

Given this, representation in science is a two-tier process. The model represents a hypothetical entity and then this hypothetical entity represents the target system. Frigg calls the first tier p-representation and the second T-representation (2010b, 100). Given the centrality of hypothetical entities in Frigg's account of models, the natural question is: what is a hypothetical *entity*? Is it something that does not exist or something that might exist but we don't yet have decisive evidence for it (e.g., the hypothetical ether or the planet Vulcan)? In the present context, it is *not* the latter.

I take it that a hypothetical entity is, typically, a putative entity that is posited for theoretical/explanatory reasons but it might not be there after all. For instance, the planet Vulcan was posited by Le Verrier in an attempt to explain the anomalous perihelion of Mercury. No such planet was observed. Hence, it was concluded that Vulcan is not there: it does not exist. Apparently, a hypothetical entity like the ones noted above is not necessarily a fiction. Vulcan might, after all, have been a real planet. To see how, let us just note that planet Neptune was posited as a hypothetical entity to explain the anomalous orbit of Uranus. It turned out that Neptune graduated from being hypothetical to being real.

All this is *not* what Frigg has in mind when he talks about models being hypothetical entities. According to Frigg, hypothetical entities should be taken not to exist. They can never graduate to reality. They are, once and for all, imagined entities. So let us substitute 'non-existent' for 'hypothetical' in Frigg's quotation noted above:

models are non-existent entities that, as a matter of fact, do not exist spatio-temporally but are nevertheless not purely mathematical or structural in that they would be physical things if they were real.

What does this say? Let's leave aside the seemingly paradoxical 'non-existent entities'. What is implied is that models don't exist but if they existed they would be in space-time. Why do models not exist? Because they don't exist in space-time; hence, because they are not concrete. Hence, the thesis above says:

Models don't exist because they are not concrete; but if they existed they would be concrete.

This thesis equates the real with the concrete. Hence, it presupposes that only concrete things do exist and *could* exist. But, as we have noted, this simply begs the question.

Frigg shares the view that models are props in games of make-believe, where the descriptions that introduce the model-system are the props. There are rules of generation too. Together they define an imaginary system; they attribute to it various properties and precisely because scientists are within this make-believe game, they are entitled to imagine that the hypothetical entity has these properties. Hence, unlike Walton, Frigg *does* posit a hypothetical entity, a fictive entity which is characterized by the model. What kind of status does this entity have? What, for instance, is one committed to when one takes the Carnot engine to be a hypothetical (imaginary) entity?

Here is what Frigg says: "What metaphysical commitments do we incur by understanding models in this way? The answer is: none. Walton's theory is antirealist in that it renounces the postulation of fictional or abstract entities, and hence a theory of scientific modeling based on this account is also free of ontological commitments" (2010b, 120). We have already seen some of the problems Walton's theory faced in its attempt to renounce commitment to entities. So it's far from clear that renouncing such commitments is viable. But in Frigg's way of putting the fictionalist case, things are worse. For we are left totally in the dark as to what models are, if they don't exist. Are they nothings? But if they are nothings, how can they have properties? Frigg notes that "saying that a hypothetical entity possesses certain properties involves nothing over and above saying that within a certain game of make-believe we are entitled to imagine the entity as having these properties" (2010b, 116). What exactly are we imagining as possessing properties if not an entity? And then the question is: how best do we conceive the ontic status of this entity? Contrary to Frigg's claim that "there is nothing mysterious about ascribing concrete properties to nonexistent things" I think it's utterly mysterious how the use of a theoretical description as a prop is not (and cannot be) about anything and yet to describe something!

Let's go back to the Carnot engine. To say that this is an imagined non-existent 'entity' fails badly to explain that this 'entity' set limits to concrete heat engines.

The reversibility of the Carnot cycle is not an imagined property of an imagined 'entity'. Rather it is real property of a real, albeit abstract, entity. To be sure, Frigg leaves it open that there might be reasons to prefer a realist view of models over a factionalist one. His point is that "whatever these reasons may be, the needs of science are not one among them (2010b, 120). That's too quick, however.

To see it let us note that science is replete with mixed statements, that is statements which compare models with actual systems. It is said, for instance, that the earth-moon system is like a two-body (Newtonian) model. Or it is said that no concrete heat engine matches in efficiency the Carnot-engine. Frigg calls these statements transfictional. If models are nonexistent entities then, as Frigg admits, transfictional propositions—apparently—involve comparing something with a nonexistent object. But how can that be? What is the truth-maker of a transfictional proposition if some alleged part of it is missing? How can we compare Y with X if there is no X?

Note that on the view that models are abstract entities there is no mystery here. The model (that is, the entity described by a literal understanding of the model description) is a real but abstract entity. Hence, what Frigg calls ‘transfictional propositions’ are not *transfictional*. They are ordinary propositions which compare two entities: one concrete and one abstract. As in the case of every comparison, there are respects and degrees of similarity and likeness. But these are fixed by the context of comparison. For instance, when we compare an ordinary steam engine with a Carnot-engine, the context abstracts away from various features of the concrete engine (e.g., that is contained in a locomotive, that it is made of steal etc) and focuses on features that are relevant to its operation qua heat-engine (e.g., that there is a temperature difference).

How can a fictionalist account for comparative (so-called transfictional) propositions? Frigg’s answer, simply put, is that there is no need to assume any object other than the concrete one. The comparison, he says, involves “properties rather than objects, which makes the original puzzle go away. Crucially, then, truth conditions for transfictional statements (in the context of scientific modeling) come down to truth conditions for comparative statements between properties, which are unproblematic in the current context (that is, the problems that attach to them have nothing to do with issues surrounding fictional discourse)” (2010b, 119).

Things, however, are not so simple. As noted already, properties need bearers and non-existent bearers have no properties. Even if we think of entities as clusters or properties without anything like a bare particular underneath, we still talk about entities! Take Frigg’s own example (coming from David Lodge’s novel *Changing Places*): “For instance, when I say ‘my friend James is just like Zapp’ I am not comparing my friend to a nonexistent person. What I am asserting is that both James and Zapp possess certain relevant properties (Zapp possesses properties in the sense explained above) and that these properties are similar in relevant ways” (2010b, 119–120). The sense “explained above” is that a fictional entity is part of a make-believe story within which it is imagined that it has various properties. But if there is no Zapp, in what sense does he possess properties like James’s? In what sense is it true to say that ‘my friend James is just like Zapp’?

The best way to make sense of this comparison is to think of it as being between a real story and a make-believe story, which is being treated as being referential and true. What makes the statement of comparison true, if any, is the fact that James possesses properties that Zapp-in-the-novel possesses too. Zapp-in-the-novel is no less an entity than James. To be sure, Zapp-in-the-novel, (better: Zapp-in–*Changing-places*) is created by David Lodge. Zapp-in-the-novel is brought into existence by David Lodge. Zapp-in-the-novel exists dependently on a novel while James exists independently of novels. Actually, Zapp-in-the-novel has only the properties assigned to him by his creator. But this does not change the fact that in transfictional claims, there is comparison between *entities*; however, one of them has existence parasitic on the existence of the novel.[4]

When it comes to models-as-fictions, things are a bit more complicated. Take, for example, the comparative statement:

The earth-moon system is like a two-body Newtonian system.

On the fictionalist view, we run into the same problem as above: how can this be true if there is no two-body Newtonian system? If, that is, the two-body Newtonian system is non-existent? Note that, unlike the case of Zapp-in-the-novel, there is no reason to think that a two-body Newtonian system has dependent existence. The chief reason for this is that a two-body Newtonian system has, as it were, a life outside Newton's theory. This is mainly because the two-body Newtonian system is the product of abstraction which cut off part of being. The product of this process is discovered and not created, since what are discovered, in the end, are the causally relevant properties of a part of being. A causally relevant property is, for instance, having mass and not the shape in which the mass of a body takes.

Hence, unless there are independent reasons to think that a two-body system is a fiction, then the statement above is semantically the same as

The trajectory of Mars is like an ellipse.

Here there is comparison between an actual trajectory and a mathematical entity. A concrete object is (successfully) compared with an abstract one. In any case, unlike ordinary transfictional statements comparing James to Zapp-in-the-novel, statements comparing models with actual systems are liable to representational success and if true, are true by means of a deep connection or similarity between the model and the concrete system.

9.5 Conclusions

The key question is: are there independent, not question-begging, reasons to take models as non-existent? As fictions? The key question is, then, metaphysical. The neo-fictionalist account of models takes the hard line that models are not part of reality. What is reality? Everything that there is. But why are models not part of reality? Models are certainly not actual, spatio-temporally located or causally active entities. They are not part of the concrete content of the natural world. But that's not reason enough to treat them as unreal. Novels are, typically, works of fiction. There is no question of them representing an independently given domain; there is no issue of success or failure (though a novel may have factual elements). This is not the case with scientific theories. Models function within scientific theories and

[4] For more discussion of the transfictional statements, see Peter Godfrey-Smith (2009).

scientific theories, unlike novels, have representational and empirical content. They are judged in terms of success or failure to connect with their intended domain.

The key motivation for claiming that models are fiction is a disdain for abstract entities. Nominalism should be behind the metaphysics of models-as-fictions. I have argued against nominalism in my (2010). So I will end with a note on the kind of realism that I advocate. I call it moderate realism: models and other abstract entities which encode physical properties, depend for their existence on concrete physical entities but they are real nonetheless.

Acknowledgement An earlier version of this paper was presented at the workshop "Scientific Contents: Fictions or Abstract Objects?", organized by the EPISTEME Research Group, at the University of Santiago de Compostela, in January 2017. Many thanks to Xavier de Donato and José L. Falguera for their kind invitation and to the participants for useful comments and questions. Thanks are due to an anonymous reader and the editors of this volume for their patience and support.

References

Aristotle. (1991). Metaphysics. In J. Barnes (Ed.), *Complete works*. Princeton: Princeton University Press.

Bäck, A. (2014). *Aristotle's theory of abstraction*. Dordrecht: Springer.

Carnot, S. (1824). *Reflections on the motive power of fire: And other papers on the second law of thermodynamics*. New York: Dover Publications.

Clapeyron, E. (1834). Mémoire sur la Puissance Motrice de la Chaleur. *Journal de l'Ecole Royale Polytechnique, XIV*, 153–190.

Duhem, P. (1892). Some reflections on the subject of physical theories. In *Essays in the history and philosophy of science* (R. Ariew and P. Barker, Trans. and ed., pp. 1–28). Indianapolis: Hackett Publishing Company, 1996.

Duhem, P. (1893). The English school and physical theories. In *Essays in the history and philosophy of science* (R. Ariew and P. Barker, Trans. and ed., pp. 50–74). Indianapolis: Hackett Publishing Company, 1996.

Forster, E. M. (1962). In O. Stallybrass (Ed.), *Aspects of the novel*. Harmondsworth: Penguin Books.

Frege, G. (1884). *The foundations of arithmetic* (J. L. Austin Trans.). Evanston: Northwestern University Press, 1980.

Frigg, R. (2010a). Models and fiction. *Synthese, 172*, 251–268.

Frigg, R. (2010b). Fiction and scientific representation. In R. Frigg & M. Hunter (Eds.), *Beyond mimesis and convention*. Dordrecht: Springer.

Godfrey-Smith, P. (2009). Models and fictions in science. *Philosophical Studies, 143*, 101–116.

Horsten, L., & Leitgeb, H. (2009). How abstraction works. In A. Hieke & H. Leitgeb (Eds.), *Reduction – Abstraction – Analysis* (pp. 217–226). Frankfurt: Ontos Press.

Mally, E. (1912). *Gegenstandstheoretische Grundlagen der Logik und Logistik*. Leipzig: Barth.

Newton, I. (1729). *The mathematical principles of natural philosophy* (A. Motte Trans.). Berkeley/Los Angeles: University of California Press, 1934.

Psillos, S. (2010). Scientific realism: Between Platonism and nominalism. *Philosophy of Science, 77*, 947–958.

Psillos, S. (2011). Living with the abstract: Realism and models. *Synthese, 190*, 3–17.

Russell, B. (1903). *Principles of mathematics*. Abingdon: Routledge (2009).

Toon, A. (2012). *Models as make-believe. Imagination, fiction and scientific representation*. Palgrave-MacMillan.

Vaihinger, H. (1911). *The philosophy of 'As If'* (C. K. Ogden, Trans.). London: Kegan Paul, Trench, Trubner and Co Ltd (1935).

Walton, K. (1990). *Mimesis as make-believe*. Cambridge, MA: Harvard University Press.

Wang, J. (2011). *Modern thermodynamics. Based on the extended Carnot theorem*. New York: Springer.

Zalta, E. (1983). *Abstract objects: An introduction to axiomatic metaphysics*. Dordrecht: D. Reidel.

Chapter 10
Models and Denotation

Fiora Salis, Roman Frigg, and James Nguyen

Abstract Many models function representationally. Considerable differences notwithstanding, most accounts of representation involve the notion that models denote their targets. Denotation is a dyadic relation that obtains between certain symbols and certain objects. This does not sit well with the fact that many models are not concrete objects. If a model does not exist, how can it denote? We present an antirealist theory of models that reconciles the notion that models don't exist with the claim that there is real denotation between models and their targets.

Keywords Models · Scientific representation · Denotation · Mental file · Intentionality · Fiction · Pretence · Truth in fiction · Exemplification

10.1 Introduction

Denotation is a dyadic relation that obtains between certain symbols and certain objects. Symbols can be of many different types, including linguistic, pictorial, and mental. Proper names are paradigmatic examples of denoting symbols. For example, we can use a token of the name "Diogenes" to denote a particular individual, namely Diogenes the Cynic.

F. Salis
Department of Philosophy, University of York, London, UK
e-mail: fiora.salis@york.ac.uk

R. Frigg (✉)
London School of Economics and Political Science, London, UK
e-mail: r.p.frigg@lse.ac.uk

J. Nguyen
Department of Philosophy, University College London, London, UK

Institute of Philosophy, School of Advanced Study, University of London, London, UK

Centre for Philosophy of Natural and Social Science, London School of Economics and Political Science, London, UK
e-mail: james.nguyen@sas.ac.uk

J. L. Falguera, C. Martínez-Vidal (eds.), *Abstract Objects*, Synthese Library 422,
https://doi.org/10.1007/978-3-030-38242-1_10

This raises two questions. The first concerns the nature of the symbols themselves: what objects are they and how do we identify them? We call this the *identification problem*. The second concerns the relation between the symbol and the object the symbol stands for: in virtue of what does it hold? We call this the *relation problem*. Philosophers of language spent a great deal of time investigating the nature of linguistic denotation,[1] with a particular focus on proper names. The focus in this endeavour was on the relation problem because philosophers of language – rightly – took the nature of linguistic symbols to be clear enough not to worry greatly about it.[2]

In recent years the issue of scientific representation has attracted considerable attention and a range of different accounts of scientific representation have been proposed (for a review see (Frigg and Nguyen 2017)). Considerable differences notwithstanding, many of these accounts rely on denotation. The structuralist conception requires that there is a correspondence between objects in the model and objects in the target, as well as between properties and relations of the model and of the target (see, for instance, French and Ladyman 1999), and Contessa (2007) explicitly construes these correspondences in terms of denotation. Advanced versions of the similarity view require hypotheses that denote the target system (Giere 2010). Accounts based on Goodman and Elgin's notion of representation-as (Elgin 2010, 2017) see denotation as the core of representation (Frigg and Nguyen 2018).

A further point of convergence is that many accounts recognise models as the units that are doing the representing: it's models that represent target systems. Hence, in the case of scientific representation models take the place of proper names in linguistic representation. This adds an additional layer of complexity to the problem. While philosophers of language can rest reasonably content that the ontology of proper names (and linguistic symbols more generally) is sufficiently unproblematic to set the identification problem aside, there are no such assurances concerning scientific models and a great deal of ink has been spilled on the problem of the ontology of scientific models in recent years.

The so-called fiction view of models submits that scientific models are, from an ontological point of view, akin to characters and places in literary fiction. As Godfrey-Smith puts it "modelers often take themselves to be describing imaginary biological populations, imaginary neural networks, or imaginary economies. [. . .] Although these imagined entities are puzzling, I suggest that at least much of the time they might be treated as similar to something that we are all familiar with, the imagined objects of literary fiction. Here I have in mind entities like Sherlock Holmes' London, and Tolkien's Middle Earth" (Godfrey-Smith 2006, 735). Combining the views that models are the vehicles of scientific representation,

[1] Or reference – in the philosophy of language the terms "denotation" and "reference" are often used as synonyms.

[2] Kaplan's (1990) investigation into the nature of words (and proper names in particular) as the media of denotation is a noteworthy exception.

that denotation lies at the heart of representation, and that models are like fictions gives the position that fictions denote target systems in the world.[3]

Previously we have said that denotation is a relation between a symbol and an object. At least *prima facie*, only existing objects can enter into relations and so one may wonder how a fictional model can possibly denote a target system. Indeed, one might take this to be a *reductio* of the position: fictions don't exist and hence can't denote anything. A position that combines the fiction view of models with an account of representation that involves denotation must therefore be incoherent.[4]

In what way exactly a fiction account of models faces this challenge depends on one's metaphysics of fiction and on the exact analysis of models. Realists about fictional entities might reply that fictions do exist and that denotation therefore poses no problem. Whether realism actually offers a quick fix to the relation problem is an interesting question, but for want of space we cannot pursue it here. Our focus is on antirealism (with respect to fictions and models), and we discuss how antirealists can respond to the incoherence charge. As such our argument in this paper addresses the conditional question: if antirealism about models is correct, then how do they denote? We will briefly return to the other options available to someone who subscribes to the position that models are fictions (and therefore not concrete objects) that nevertheless denote their targets in the conclusion.

Throughout the paper we are concerned with the identification problem. The identification problem is conceptually prior to the relation problem because one can ask *how* something denotes only once it is clear that the "something" in question is the kind of thing that could at least in principle enter into a denotation relation. As we have just seen, in the context of the fiction view this cannot be taken for granted. The aim of this paper is to show how the fiction view can give a positive answer to the identification problem even in the context of an antirealist view of fiction, thereby paving the ground for a future discussion of the relation problem.

We discuss the identification problem in the context of modelling in the natural sciences. The kind of models we have in mind are the billiard ball model of gas, the Newtonian model of the solar system and the Lotka-Volterra model of predator-prey interaction. The term "model" is sometimes used with a different meaning in normative contexts, where models are used to describe fictional characters as representatives of vice and virtues or right and wrong. This is also relevant in the context of using (at least some) models from the social sciences, e.g. models of decision-making, which have normative rather than descriptive content (i.e. the model doesn't represent what an agent does, it represents what an agent *should*

[3]Fiction accounts of models have also been advocated by Barberousse and Ludwig (2009), Frigg (2010), Frigg and Nguyen (2016), Giere (2010, 278) (although he stresses that this is restricted to ontology, functionally models and fictions might come apart), Salis (2019) and Salis and Frigg (2020). Levy (2015) and Toon (2012) present accounts that appeal to fiction, but are designed in way that does not require that a model denote a target. For a discussion of their account see Frigg and Ngueyn (2017, 86–88).

[4]The objection actually applies more generally: anyone who thinks that models are not concrete objects but denote their targets will face the same challenge.

do). Such a normative use of models, and representations more generally, is an underexplored topic in the literature on scientific representation. However, this is not the place to fill that lacuna: irrespective of the particular issues that arise from their normative content, presumably they still need to be identified, and still denote their targets. Thus, any account of them will also require a resolution of the issues that we are addressing here. As such, we set aside the thorny issue of how to analyse such normative contexts here.

The structure of the paper is as follows. In Sect. 10.2 we introduce an antirealist position about scientific models that builds on Walton's (1990) account of fiction. In Sect. 10.3 we draw on the literature in philosophy of mind and psychology to provide an answer to the question of how we can imagine things about non-existent objects. In Sect. 10.4 we apply this to scientific models and argue that by identifying models with their descriptions and the fictional truths we are prescribed to imagine when engaging with them, we can accommodate the idea that models can denote their targets. Section 10.5 discusses some ramifications of this position, and Sect. 10.6 integrates it with the DEKI account of scientific representation (Frigg and Nguyen 2016, 2018). Section 10.7 concludes.

10.2 Antirealism About Models

Antirealists claim that there are no fictional objects. Walton (1990) developed a paradigmatic antirealist theory of fiction as a game of make-believe that has been influential for contemporary developments of the fiction view of models.[5] Construing models as akin to works of fiction and games of make-believe does not amount to portraying modelling as something unserious or even frivolous. Walton emphasises that one of the reasons why children engage in games of make-believe is to cope with their environment. He mentions an extreme case of a game played in Auschwitz called "going to the gas chamber"[6] and argues that the children playing the game were "facing the reality of genocide with the utmost seriousness" (1990, 12). The ability to engage in games of make-believe continues in our adult life when we interact with works of art. We imaginatively engage with literary fictions, dramas and other works of art that can in no way be dismissed as unserious (think of *Anna Karenina* or *Othello*). Scientists imaginatively engage with models for many serious purposes, including learning about reality.

Walton argues that works of fiction are akin to games of make-believe, which he characterises as imaginative activities involving props. Props are objects like toy trucks or texts of works of fiction. Props make propositions fictionally true in virtue of certain prescriptions to imagine that are stipulated, or implicitly understood, to be in force within a game. Fictional truth is a property of imagined propositions:

[5] See, for instance, Frigg (2010), Levy (2015), and Toon (2012).

[6] See Opie and Opie (1969) for a description of this game.

fictionally true propositions are those that *ought to* be imagined in the relevant game. Fictional truths divide into two kinds, primary fictional truths and implied fictional truths. The former are generated directly from the text of a fictional story, while the latter are generated indirectly from the primary fictional truths via principles of generation. These principles might vary from case to case, depending, for example, on the genre of the fiction. Walton discusses the reality principle, which keeps the world of the fiction as close as possible to the real world, and the mutual-belief principle, which is directed toward the mutual beliefs of the members of the community in which the story originated. Games of make-believe are of two main kinds, authorised and unofficial. They are authorised when they are constrained by the author's prescriptions to imagine. They are unofficial when the principles of generation constraining them are *ad hoc*. Finally, games of make-believe can involve imaginings that are about real objects (we can imagine that Churchill was a member of the communist party) as well as about fictional objects (we imagine that Sherlock Holmes plays the violin). But either way imaginings do not have any ontological import and do not commit us to postulating fictional entities.

Applying these ideas to scientific models gives us the following picture. Models involve model descriptions that function as props. They are descriptions of models we find in scientific papers and textbooks. They prescribe imagining certain fictional truths. The model content includes the explicit fictional truths prescribed by the model description and the implied fictional truths generated by certain principles of generation that are in operation in the context in which the model is used (these can be, for instance, assumed to be laws of nature or other general principles of the scientific field in which the modelling practice is embedded). Typically, model descriptions express general propositions about properties and relations in virtue of having those properties and relations among their constituents. For example, the proposition "all humans are mortal" is about the properties of being human and being mortal. Model descriptions can, but often do not, describe objects in the world. Indeed, there need be no objects that satisfy the descriptive conditions stipulated by scientists.[7]

Nevertheless, model descriptions *seem* to prescribe imaginings about some particular fictional systems. Thomson-Jones (2010) emphasises this aspect of the phenomenology of the modelling practice – what he calls the *face value practice* – when he says that a model description "has the surface appearance of an accurate

[7] Walton assumes referentialism, the position that utterances of sentences containing proper names express singular propositions. For instance, the proposition "Saint Paul's Cathedral is Northern Europe's biggest church" is directly about St Paul's in virtue of having St Paul's among its constituents. This view entails that utterances of sentences containing fictional names (names without referents) express either no proposition or a gappy proposition. Walton (1990, Ch. 10) assumes the former and argues that utterances of sentences containing fictional names are to be analysed in terms of *kinds of pretence*. Friend (2011) and Salis (2013) emphasise that this is insufficient to distinguish different kinds of pretence that seem to be about different fictional objects. They offer alternative analyses in terms of gappy propositions and participation in different networks of information (Friend 2011) and different name-using practices (Salis 2013).

description of an actual, concrete system (or kind of system) from the domain of inquiry" (2010, 284). Fibonacci's model of population growth, for instance, seems to prescribe imaginings about a particular fictional population of rabbits. Furthermore, this model system is identified with the vehicle of the representation relation between the model and the world, and so this model system *seems* to denote the target of the model.[8]

Hence, antirealists actually face two problems. The first is to make sense of the face value practice that takes model descriptions to be about a particular model system even though there are no such systems. We call this the *problem of model systems*. The second is the problem concerning denotation that we have already mentioned: how can models be the vehicles of denotation if they don't exist? It turns out that a reflection on the first problem also offers a solution to the second problem, and so we start with a discussion of the face value practice.

10.3 Object-Directed Thoughts and Mental Files

To address the face value practice, it pays to note that the problem of model systems is an instance of the more general problem of the object-directedness – or *intentionality* – of mental states. To get a grip on this problem we now introduce a cognitive account of the intentionality of thoughts that are directed to particular objects, and then show how the account offers an answer to the problem.

The account is best introduced with an example. The philosophy department has to move to a new building. The head of department receives a dossier about the building, containing a detailed description of the layout of the rooms along with architectural plans. But she can't inspect the building because it is undergoing extensive refurbishment and the administration considers it unsafe for her to visit the building site. So she forms a view of the building and starts planning the move of the department solely based on the content of the dossier.

Let us have a closer look at what happens in this process. The *dossier D* is a conglomerate of sentences in English, drawings, plans, etc. The dossier has *content C*. The content is objective and publicly accessible to everybody who is able to read *D*. The content of the dossier is *about* the building *B*. When the head of department reads the dossier she forms a *mental file F* of the building.[9] Mental files are modes

[8]In fact, Weisberg (2007) identifies the existence of a secondary object that does the representing as a defining feature of modelling.

[9]In the late1960s philosophers of mind and language introduced the notion of a mental file as a cognitive representation of concrete objects as individuals rather than as the possessors of properties. Originally, Grice (1969, 141–142) introduced this notion under the label "dossier" in his discussion of vacuous names and referentially used descriptions. The idea is that our thoughts latch onto reality in a direct way, i.e. through a perceptual relation with individual objects rather than through the mediation of a descriptive condition that looks for the object as the satisfier of a certain set of qualitative features. In line with this idea, Perry (1980) introduced the term "mental

of presentation of individual objects. They involve information about properties that one takes the object to have. So, the head of department's mental file contains information, which she takes to be about the building. F is informed by C (recall that the only source of information about B is D), but it need not be identical with it. In fact, F is how the object is given to the head of department: F contains her personal construal of the building, which can, but need not (and usually does not), line up with C; and, indeed, different people can (and usually will) have different mental files about the same object. The head of department may not have realised that there is a storage room in the basement; she may have paid no attention to the roof structure; and she may have miscounted the number of offices on the first floor. Her mental file differs from C in all these respects.

Information contained in F can be construed as a list of predicates that one takes as satisfied by the object (Recanati 2012, 37). Predicates can be relational, and importantly, may involve other mental files. As a consequence, two files could appear in each other's list.[10] To avoid a regress, one cannot construe mental files as constituted and identified by their predicates and corresponding properties. The properties are merely associated with the file, and information can be added to and deleted from the file without changing the file itself. And, as

file" to account for the phenomenon of continued belief and he appealed to the same notion to account for the phenomenon of co-reference in his (2001, 128–146). Bach (1987, Ch. 3, spec. 34–39, 44) deployed mental files in his discussion of *de re* thought. More recently, Jeshion (2010) presented a new theory of singular thought as thought from mental files. Friend (2011, 2014) appeals to mental files to explain the phenomenon of intersubjective identification of fictional characters within fictional antirealism. Intersubjective identification of the same object, or co-identification, is further explained in terms of participation in the same information network (Friend 2011, 2014) or the same name-using practice (Salis 2013) supporting the mental files. Linguists have used the notion of a mental file as discourse referents (Heim 1982; Kamp and Reyle 1993; Karttunen 1976). Cognitive psychologists have introduced the analogous notions of *object files* to study visual representations in adults' object-directed attention (Pylyshyn 2000, 2001, 2007; Fodor and Pylyshyn 2014) and *object concepts* to theorise about object representations in infancy (Spelke 1990; Baillargeon 1995; Carey 2009). Pylyshyn (2001, 129) draws an explicit connection between the philosophical literature on mental files and the notion of object files to emphasize the purely causal relation between object files and their referents. Philosophers Murez and Recanati (2016) make some important distinctions between Pylyshyn's notion of object files and mental files by underlying the conceptual nature of the latter. They emphasise that mental files can store qualitative information about their objects (and in this sense they can be construed as conceptual representations of individual objects). However, this qualitative information is not used to fix the referent of the mental file. Information can be updated, retrieved and deleted without changing the referent of the file. It is in this sense that we say that mental files represent concrete objects as individuals rather than as the possessors of properties. Perner et al. (2015) explicitly appeal to mental files to develop a cognitive theory of belief representation in infancy.

[10]This happens, for instance, when a file involves an expression that *appears* to be a proper name. The head of department says "room 425 is too small to host the admin office". A singular term like "Room 425" has its own individual mental file, and so does each item described in the dossier. In fact, there can be a hierarchy of files. But the files contain only information (predicates) about the objects. They don't contain singular terms. They are cognitive representations in the mind that stand for objects in reality, if there are any. They are associated with singular terms without including them.

noted previously, information can be subjective and idiosyncratic to the extent that different individuals can associate different information with their mental files for the same object independently of whether this exists or not.

Mental files are also modes of presentation of particular objects, and so they play cognitive roles akin to Fregean senses. This solves both the problem of cognitive significance and the problem of object-directed but objectless thoughts. The former is the problem of explaining how one can have different thoughts about the same object, possibly even without realizing that one is thinking about one and the same object. The solution in terms of mental files is that one can associate two distinct mental files involving different information with the same object, analogous to the way in which Frege's introduction of senses accounts for the classical example with the morning star and the evening star. With "Phosphorus" and "Hesperus" associated with "the morning star" and "the evening star" respectively, we can explain their different cognitive significance by claiming that the belief that Phosphorus will rise and the belief that Hesperus will rise are associated with different mental files involving different information.[11] The account we have introduced is summarised in Fig. 10.1.

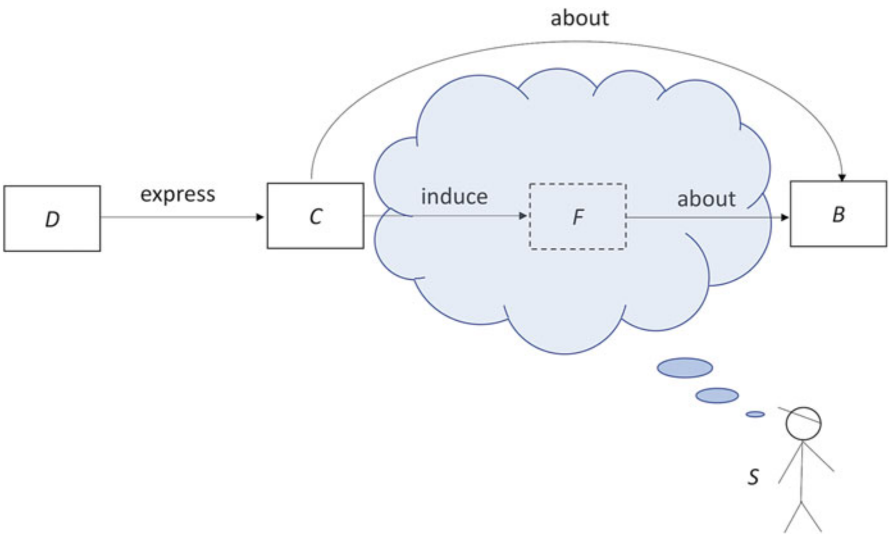


Fig. 10.1 Mental file of an existing object

[11]Fregean senses are usually interpreted as descriptive modes of presentation of objects, i.e. descriptive conditions such as "the morning star", that are parts of the propositional content of thoughts. For example, the content of the thought that "Phosphorus will rise in the morning" will be that *the morning star* will rise in the morning. Mental files, however, are not necessarily interpreted as descriptive modes of presentation that enter into the propositional content of thoughts. See, e.g., Recanati (2012) for a critical discussion of the relation between mental files and descriptivism.

Now let us change the story slightly. Rather than moving into a pre-existing building the university decides to commission a new building for the philosophy department. In this case the dossier that the head of department sees does not contain information about an existing building. It contains information about something that does not exist at all: if the building has not yet been built there is no *B* and on pain of incoherence neither the *C* nor *F* can be about *B* (at least if we take it that *B* must exist to be the subject of an intentional thought). So, strictly speaking, *C* and *F* are not about anything. Nevertheless, we habitually engage into a practice of reflecting and talking "about" the inexistent "building" as if it was real. The head of department may object that the "building" does not have enough offices and that the seminar room is too small. How can one make sense of this practice if there is no building?

Realists about intentionality argue that every mental state needs an object of some kind and therefore argue that there exists a fictional or abstract object – the future building – that the discourse is about. Antirealists about intentionality disagree and submit that not every mental state needs an object. Our focus here is on imagination. Imagining a pink dragon (which is a kind of mental state) does not require that us to postulate that there is a pink dragon that we are thinking about, and, likewise, imagining the future philosophy building does not require the postulation of a fictional or abstract object that we are thinking about.[12] Regardless of what antirealists say about other mental states, imaginings at least are not necessarily directed at anything in the world.

But now we're faced with a generalised version of Thomson-Jones' face value practice: we habitually talk about pink dragons and future buildings as if they were ordinary objects and yet there are no such objects. But how can it seem to us that we are thinking about an object when there is no object to think about? At this point mental files come to rescue because they explain the seemingly oxymoronic phenomenon of object-directed yet objectless thoughts.

When thinking about pink dragons or inexistent buildings we deploy a mental file. As we have seen, a mental file is akin to a concept or a mental representation that stands in for an individual object. Yet, as a cognitive structure for the storage of information, mental files incur no ontological commitments, and a subject can (and usually does) have mental files both for real objects and figments of the imagination. Someone may have a mental file for "hotel" as well as for "griffin", where the file simply is a list of properties that the subject takes the object to have. Since mental files are standardly associated with something that exists, it *seems* to us that we are thinking about an object *whenever* we deploy a mental file – even if there actually is no object. This is because thoughts that seem to be about an object without there being one effectively engage the same types of cognitive resources as thoughts about existing things. From a cognitive point of view, internally, there is no genuine difference between the two cases.

[12] In this section we use "pretend" and "imagine" in their non-technical sense, which is broader than the technical sense introduced in Sect. 10.2.

For this reason, when we deploy a mental file that is not about an object, it is natural to think and talk as if we were thinking and talking about an object even if we know full well that there isn't one. This cognitive illusion is practically impossible to escape. It is an effect of the deployment of the same cognitive structures that originally enable our thoughts and discourse about real objects. This is why we think and talk *as if* there were an object. One can then say that these thoughts seem to be about an imaginary object toward which the mental file is directed. But actually, this is only pretend-aboutness rather than genuine aboutness. What this means is that there seems to be an imaginary object, one that (fictionally) exists in our minds. Effectively, this seeming object is a construct of the imagination. But, as we said before, imagining that something is so and so does not commit us to postulating that there really is something. In pretence, we can manipulate, explore, and transform this imaginary object just like we would manipulate, explore, and transform a concrete object. What this actually means, however, is that we can only pretend to manipulate, explore, and transform the imaginary object. What we really do is updating, adding, or deleting information from the mental file for the imaginary object.

When the head of department talks about the new philosophy building, her mental file is not about a building (it's yet to be built!), but she, as well as her interlocutors, pretend that the information they have (the one that is stored in the mental file) is about the same particular object. They do so by engaging in a way of thinking and talking that seems to involve reference to an object, and the content of their mental files is about this imaginary object. Ultimately, what they are really doing is not thinking and talking about a real object (indeed, they are even aware of this). Rather, they manipulate, explore, and (possibly) transform the information they have to better plan for the future building. The head of department might add comments to the dossier and ask for clarifications, further information, and even changes to the plan. Her interlocutors will understand these comments as being about the future building (the imaginary object), and they will clarify, explain, and amend the dossier by changing its content. Once the content has been changed, the mental files that the head of department and her interlocutors have will also change in a way that reflects the agreed content of the new dossier. They will update, add, or delete information that they take to be, in the imagination, about the same object. And as a consequence, the imaginary object itself will be thought about and described (in the imagination) as being different from the way it originally was thought about and described. The account we introduced is summarised in Fig. 10.2.

10.4 Revisiting Models

The account of object-directed thought that we introduced in the last section equally applies to models if we associate *D* with the model description (which, like the architectural dossier, can contain a mix of verbal descriptions, drawings, graphs, etc.); *C* with the "content" of the model (which, in the Waltonian framework we're working in, includes the primary fictional truths from *D* and also the secondary

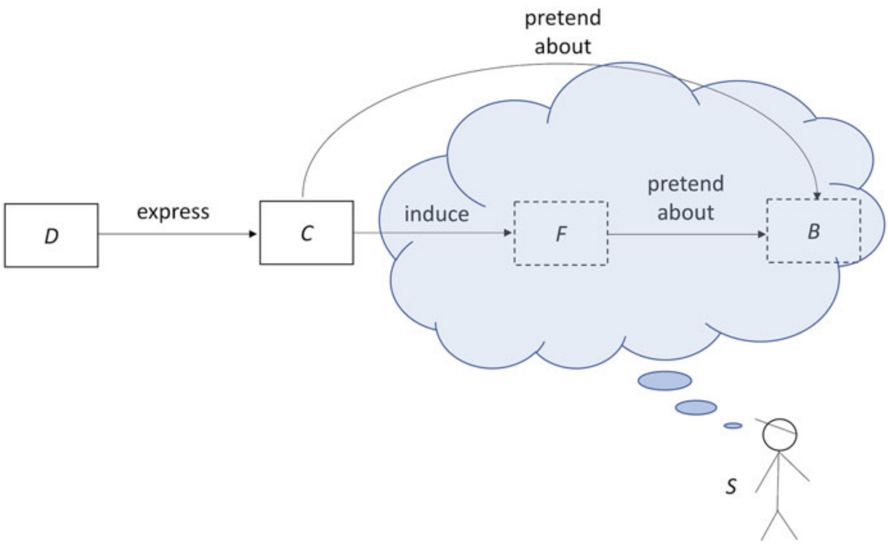


Fig. 10.2 Mental file of non-existing object

fictional truths that result when applying the principles of generation in operation); *F* with the mental file the agent forms based on *C*; and *B* with the model system. To make symbols intuitive we use "*MS*" (model system) rather than "*B*" in the context of modelling. If the model system is a material object, then we're in the situation of Fig. 10.1. The use of model organisms in biology, water-based "dumb holes" in cosmology, and the Phillips-Newlyn machine in economics are of this kind. If the model system is fictional, then we're in the situation of Fig. 10.2. Newton's model of planetary motion or Fibonacci's model of population growth belong to this group.

When engaging in the face value practice, we talk about Newton's planets or Fibonacci's rabbits as if they were actual concrete systems in the domain of inquiry. As we have seen at the end of the previous section, we can do this because we deploy a mental file. Let us illustrate this with the example of a Newtonian two-particle model system, which prescribes imagining that two homogeneous perfect spheres gravitationally interact with each other and nothing else. The model description expresses what seem to be particularised propositions, propositions that involve the properties specified by the model description and that seem to be about some particular system. But there is no system that the relevant propositions are about. The modeller's impression that she is dealing with a system is explained in terms of her deploying a particular mental file to store descriptive information that she takes to be associated with a particular system without there being any object this information is about. In this way the use of a mental file explains why it seems to us that our thoughts are about a particular system even though there is no such system.

An analysis of the face value practice requires us to introduce mental files, so that the practice now has three elements: the model description (the set of linguistic and mathematical symbols presented in papers and books), the description's propositional content, and the mental file that a scientist associates with them (which contains information that the scientist, internally, takes to be about some object independently of whether it exists or not).

Let us now turn to the second problem we identified at the end of Sect. 10.2. Often scientists think and talk as if models were the vehicles of denotation of real-world targets. They say things like “the two-particle system represents the Sun and Earth system”, or “the infinite rabbit population represents the real rabbit population”. Since, as noted in the Introduction, representation involves denotation, this claim implies that models denote their targets. How can an antirealist about models make sense of this?

A first option is to bite the bullet and say that, appearances notwithstanding, there is no denotation. When using a model to seemingly denote a target, the scientist actually claims *in pretence* that the model system denotes the target. So, models have *pretend denotation* but no “real” denotation. This does not mean that no true claims about models and their relations to targets can be made. Claims like “the Newtonian two-body system denotes the Sun-Earth system” can be analysed as being implicitly prefixed with a fictional operator, so that a full version of the claim would be something like “in the game of make-believe for Newton’s model, a scientist uses the two-particle model system to denote the sun-earth system”. This statement can be assessed for genuine truth if there is a rule according to which, in the game of make-believe for Newton’s model, a scientist can use the two-particle model system to denote the Sun-Earth system. But even if true, models have no real denotation.

However, one may think that pretend denotation is insufficient to accommodate the role the relation is supposed to play in scientific representation and that we need real denotation between models and their targets. But if models don’t exist, how could that be? It is not clear that this objection cannot be met, and we want to leave the option of building an account of scientific representation on the notion of pretend-denotation a live option. But we do not have the space to do that here. Instead we will investigate whether a solution can be offered which accommodates proper denotation.

The key to such a solution is to reconceptualise what a model is. Rather than focusing on the model system we can turn to the model descriptions and provide a new and different antirealist proposal. We can identify a model with a complex object composed of a model description D together with the model content C: $M = [D, C]$. D is the description of a model one finds in a paper or textbook. The content C is the set of both the primary fictional truths specified by the prescriptions to imagine specified by D and the implied truths derived from the principles of generation in force in the relevant context. A scientist S then uses D and C to generate, in the imagination, the model-system MS, but without there ever being any model system that is the object of S’s imaginings.

S constructs and develops the model through an act of pretence wherein certain linguistic and formulaic symbols (contained in D) are used to prescribe certain imaginings that S assumes to be about some particular system, the model system, without there being any such system. These imaginings specify descriptive information (propositions) that *S*, internally, takes to be about one and the same system and therefore stores in the same mental file. Of course, the mental file is not part of the model *M*. The mental file is a cognitive structure for the storage of descriptive information that *S* draws upon when constructing and developing the model. From a cognitive point of view, at the level of thought, this is akin to the way in which *S* would store information that she takes to be about any real individual object. And this is fundamental for the explanation of the phenomenology of the modelling practice, as we have previously seen.

In this way antirealists about model systems can preserve the indirect view without committing to exotic entities. *S* plays a game of pretence wherein certain information that is relevant for deriving the model's outcomes are stored in a mental file for an imagined system without any ontological commitment to the existence of any such system.

The upshot as regards denotation is as follows. As we have seen, the model *M* is the complex entity that is composed of the model description *D* together with the content *C*. The model *thus defined* exists and hence can enter into a denotation relation. The problem with the original version of fictionalism is avoided. This new version of antirealist fictionalism identifies the vehicle of denotation with objects that are akin to fictional stories rather than fictional characters. The model involves props that are analogous to the texts of literary fictions to the extent that they express certain propositional contents and that they prescribe imaginings that certain objects are so and so without any ontological commitment to the existence of such objects. Antirealism recognises that scientists construct and develop models in pretence, but it also allows them to genuinely use models to denote their targets by recognising that they are bona fide vehicles of denotation, and these are the model descriptions together with their descriptive contents. Model descriptions themselves do not denote any real targets just like fictional texts do not denote any real objects. Furthermore, they do not prescribe imaginings about any real systems, but rather prescribe imaginings about fictional systems without there being any such systems. In this sense, this antirealist interpretation remains an *indirect* version of fictionalism about scientific models. Model descriptions together with their contents can be construed as denotating vehicles that, under certain uses, can stand in genuine (rather than pretend) denotation relations with real world targets. This view is summed up in Fig. 10.3.

On this interpretation, models exist. They are constituted by model descriptions (linguistic and mathematical symbols) and propositional content. Both model descriptions and propositional content exist, and so the *whole* model exists. Model systems, however, don't exist (just like the fictional characters specified in fictional stories don't exist). Indeed, they are not part of the model at all. The intuition that model descriptions prescribe imagining propositions that seem to be about some particular model systems can be explained in terms of the deployment of mental files for the storage of information that they take to be about some particular object independently of whether the object exists or not. Mental files, however, are associated with the model without being part of it. They are the psychological

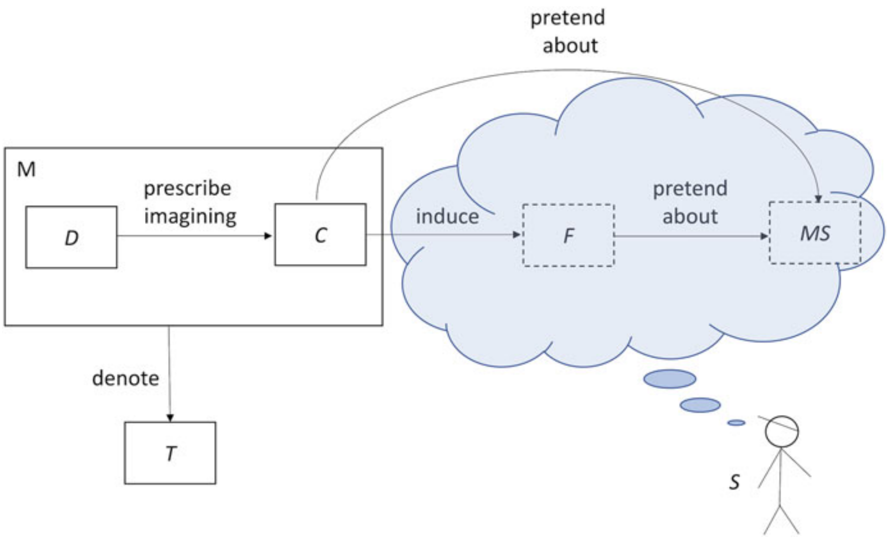


Fig. 10.3 Models

components that explain the phenomenology of the modelling practice without being part of the intersubjective, objective and public objects that are the models.

In sum, our brand of antirealism identifies models with model descriptions and their propositional content. Since both model descriptions and their content exist, there can be genuine reference to real target systems on this version of antirealism about model systems.

10.5 Reverberations

The proposal in the last section regards descriptions as a part of a model. Isn't it thereby repeating well-known mistakes? The so-called syntactic view of theories takes theories to be sets of sentences. Opponents of this view have long argued that it has the absurd consequence that every small change in the description (for instance replacing one symbol by another) results in a new theory (Suppe 2000). Does one not run into the same problem for models if descriptions are a definitional part of models?

If this was a real problem, then one would. But it isn't. In fact, a number of writers have pointed out that the syntactic view of theories is not committed to an identity criterion based on descriptions. The root of the problem is that this criterion conflates a theory and a theory formulation. A *theory formulation* is given in a particular language. It is what we encounter when we read a textbook or a scientific paper. The theory is *expressed* by a set of sentences that constitutes a formulation, and for two theories to be equivalent it isn't necessary for the theory formulations to be syntactically identical. Of course, this invites the question under what conditions theory formulations express the same theory. This question is beyond our scope here, but it is worth pointing out that there are options available. For example, Hendry and Psillos (2007) argue that two theory formulations are equivalent iff they have identical truth conditions, and Thomson-Jones (2012) makes a similar suggestion in the context of models. Quine (1975) suggests that two formulations express the same theory iff they are empirically equivalent and if the two formulations can be rendered logically equivalent by switching predicates in one of them. Relatedly, Glymour (1971) proposes that two formulations are equivalent iff they are definitionally equivalent (see also Worrall's (1984), and for a recent discussion of Quine and Glymour's proposals Barrett and Halvorson's (2016)). More recent suggestions have emerged following Halvorson's (2012) discussion of theoretical equivalence in the context of the syntatic and semantic view of theories. How each of these criteria could be utilised to account for syntatically different but nevertheless equivalent model descriptions is a question worthy of further research. Our point here is only that this is a live option, and that considering model descriptions to be part of models does not *ipso facto* imply a commitment to absurd identity conditions.

An important part of the face value practice is that we attribute properties to model systems, deem claims about model systems true or false, and investigate a model to find out about its properties (Frigg 2010). What sense can we make of this aspect of the practice if there are no model systems? Content is the key to the explanation of these aspects of the modelling practice. The attribution of properties to model systems can be explained in terms of the fictional predication of properties as expressed in the relevant propositions. Attributing the properties having-limitless-food-supplies and immortality to the imaginary rabbit population in Fibonacci's model simply amounts to imagining that the imaginary rabbits have limitless supplies of food and that they (the same imaginary rabbits) never die. We assume, in the imagination, that these properties are satisfied by the rabbit population and store the predicates in a mental file without thereby committing to the existence of any such population. These propositions will be true in the model – or f-true – if (as we stated in Sect. 10.2) they are among the prescriptions to imagine in force in the game, and they will be false otherwise. According to Fibonacci's model, the rabbits have limitless supplies of food and they never die. We can imagine that they do die, but this is false in the model and therefore it is not part of its content. In these cases, S's mental file contains predicates contradicting the predicates contained in C, just as the head of department's mental file contradicted information contained in the dossier when she miscounted the number of offices.

Moreover, we construct and develop models for the purpose of eliciting new information about the model system. This aboutness, as we explained previously, is merely pretend aboutness. What we really do is construct and develop the model description to explore its content and reveal what is implicit in it and/or increase the content by generating further claims through the principles of generation. Fibonacci's model increases its content through the implementation of some simple mathematical calculations according to which the number of rabbit pairs (one male and one female) at some time t is the sum of the numbers of pairs at the previous two times, namely $N(t_i) = N(t_{i-1}) + N(t_{i-2})$. Thus, when we imaginatively engage with a model it seems that we construct and develop a model system. But what we really do is construct and develop model descriptions that prescribe imagining certain propositions and, through the principles of generation, we further explore and expand this content. Since the model content is normatively and objectively constrained by the model descriptions and the principles of generation we can talk about right and wrong in the predication of properties of the model and in the generation of the model's outcome.

Finally, the conjunction of the views that a model contains a description and that a model denotes could be understood as implying a descriptivist theory of denotation. This is not so. The view here is that the model is a complex entity consisting of a description and its content, and it is this entity *as a whole* that denotes. In the case of the Newtonian model of the solar system, it is the entire description and its content that denotes the solar system. In some cases, a model's denotation indeed derives from the denotation of some of the terms in the description. This is what happens in the case of the Newtonian model when the use of the terms "sun" and "earth" in the model description fixes the model's denotation. However, this does not commit the view to a descriptive theory of denotation. First, the view is silent about where terms like "sun" get their denotation from and it is a matter of indifference to the fiction view of models whether descriptivism or the direct reference theory is favoured. Second, some models' denotation does not derive from the model description and comes from outside the model, as it were. As an example, consider epidemiological models as used by the crime fighters.[13] The models are about how diseases like tuberculosis spread, and accordingly the model descriptions contain terms like "disease", "infectious", etc. These models have been studied for many years by epidemiologists who investigate the dynamics of the spreading of infectious diseases. The same models are now used by the police in Chicago to predict the spreading of violent crime. So, a disease model is taken to refer to violence, but nothing in the model description is responsible for that. So, what we have said here does not entail a descriptivist answer to the relation question raised in the introduction above.

10.6 DEKI

We now want to integrate the above insights into the DEKI account of scientific representation, which is named after its four crucial aspects: models *denote* their targets, *exemplify* certain properties, which are translated via a *key*, into properties to be *imputed* to their target. The account was originally introduced in terms of concrete models (Frigg and Nguyen 2018), and previous discussions about how

[13] See Slutkin's (2016). Thanks to David Kinney for telling us about this case.

it works in the context of fictional models remained silent about how denotation worked (Frigg and Nguyen 2016, 239–240). If the account is based on the notion of a model system as an imaginary object (as, for instance, in (Frigg 2010)), then, as we have seen in Sect. 10.4, the relation between model and target is pretend-denotation. In order to accommodate the conjunction of an antirealist ontology about model systems and the view that a model ought to have "real" denotation, we had to identify a model with a description and its content. It's now time to address the ramifications this has on the rest of the account.

Having already addressed how models denote, it's important to establish how they can exemplify certain properties. Frigg and Nguyen, drawing on the work of Nelson Goodman and Catherine Z. Elgin, define exemplification as follows: X exemplifies P in a certain context iff X instantiates P and the research context highlights P, where a property is highlighted iff it is identified in the context as relevant and is epistemically accessible to users of X (Frigg and Nguyen 2016, 227). An object that exemplifies a property is an *exemplar*. A simple example is a sample card which instantiates a certain shade of red. The card both instantiates *red*, and *red* is highlighted as relevant and epistemically accessible in the context of using it in a paint shop.

Now the problem that arises is that models, identified with descriptions and their content, don't instantiate the sorts of properties that we take them to exemplify. A description of a population of rabbits doesn't instantiate *growing according to a Fibonacci sequence*, simply because it is the wrong kind of object to instantiate such a property – it consists of marks on a page and the content thus expressed. The solution to this worry is to appeal to the Waltonian apparatus: although the model doesn't, strictly speaking, instantiate the property, it is the case that "the population grows according to a Fibonacci sequence" is a secondary truth in the relevant game of make-believe. So, whilst models don't instantiate the relevant properties in the way in which a paint sample instantiates a certain colour, these properties are part of the model content C_M. C_M provides a suite of properties, and the context specifies which of these properties are highlighted as relevant. This suffices to play the role that exemplification was introduced to play. This provides us with two ways of speaking. We can talk about the model-system *MS* as pretend exemplifying certain properties. Or we can talk about the content of the model C_M containing and highlighting those properties. In the former case we are working within the scope of the pretence operator: *MS* exemplifies $P_1, \ldots P_\mathrm{n}$ in the game of make believe. In the latter, outside it: the content of the model contains and highlights (together with the research context) $P_1, \ldots P_\mathrm{n}$. These are related because the latter is true precisely because the game of make-believe prescribes us to imagine $P_1, \ldots P_\mathrm{n}$, and the context highlights those properties as relevant.

With this qualification at hand, we can now have a closer look at the model description and its content. We said that a model M is a pair consisting of a description and its content. Previously D was understood to be the *entire* model description and the content of D was the entire model content. In the context of DEKI it is helpful to break the model description up in several parts to make transparent the various elements in modelling. We now take D to be the model description narrowly construed; i.e. a description of the model system itself. This is separate from a statement of the rules of generation G which we take to be in operation in the model. G contains the statements, laws of nature or other general rules, that are at work in the model and generates secondary truths from the model's primary truths. The primary truths are expressed in D. As noted elsewhere (Frigg and Nguyen 2016, 238), D can be further subdivided into D_X and D_I. D_X is the part of the description that – in pretence – generates the "model object" (for instance two perfect spheres attracting each other with a force proportional to 1/r^2). D_I specifies how this "object" should be interpreted (for instance by instructing us to imagine the larger sphere as the sun, the smaller sphere as the earth, and the force as gravity). Each part of the description contributes to the model's content. The total content of the model C_M is therefore generated by D and G together: it contains all primary truths generated by D and all secondary truths can be derived from them using G.

In the original formulation of DEKI, a Z-representation is defined as an object under an interpretation, and a model is said to be a Z-representation where the object is used as base of the representation in a certain situation (Frigg and Nguyen 2018, 213). In the fictional case D_X generates the content that plays the role of the model object; D_I provides the interpretation of this "object"; and G generates the rest of the claims that are true of the object. Taken together D_X, D_I, and G therefore generate the content of the Z-representation. But, as we have previously seen, they also generate the model content. It follows that the model content C_M simply is the content of the Z-representation. If we let C_{ZR} denote the content of a Z-representation, then we can write: $C_M = C_{ZR}$. This equation is the "fiction equivalent" of the association of a Z-representation with an object under an interpretation in the case of a material model. C_M (or C_{ZR}) contains all the properties that the Z-representation has. Some of these are highlighted by context (and are epistemically accessible). The set of the properties so highlighted corresponds to the exemplified properties $P_1, \ldots P_{\mathrm{n}}$.

In DEKI, the exemplified properties are then "keyed up" with properties to be imputed to the model's target. What we have said so far about identifying models with their description and content is orthogonal to how both the key and imputation steps work. All that keying up requires is a collection of exemplified properties and relations (which are given by the content in combination with the context), and then the key can be introduced to translate these properties to those to be imputed to the target. For example, a model consisting of Fibonacci's equations, combined with their content in a given game of make believe (using mathematical derivations to generate secondary truths and so on), provides us with some content C. In a given context this provides us with the property of *growing according to a Fibonacci sequence* as highlighted, and this property can be translated into a property like *growing in a non-linear way* to be imputed onto the target system. The introduction of the key into the account is to accommodate the mismatches between model properties and what scientists in practise take the models to tell us about their targets,

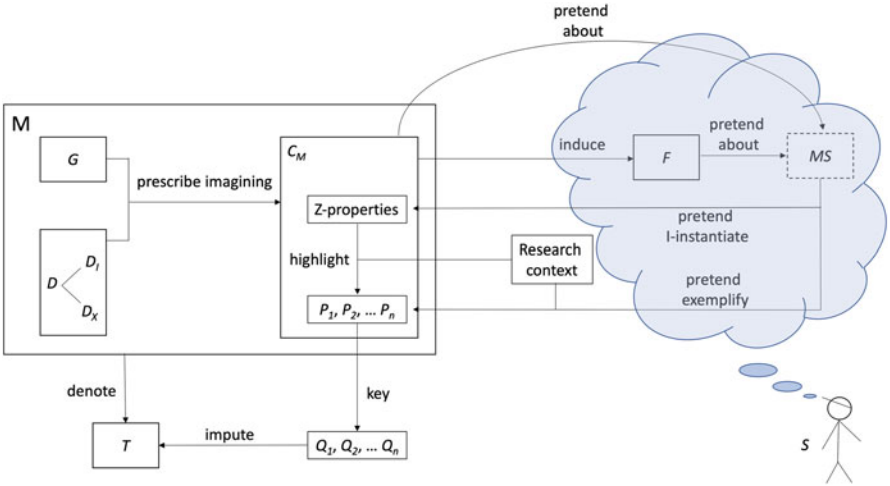


Fig. 10.4 DEKI

and once we have those properties the rest of the DEKI account carries over without further alterations.[14]

An actual model user has a mental file *F,* which is supposed to match C_M, but allows for mistakes to accommodate cases where the model user fails to investigate the model appropriately. Whilst there is no system which C_M (or *F*) is about, in the game of make-believe we imagine a model system *MS* which is their object. *MS* contains the elements of an "object" (two perfect spheres) and an interpretation (the large sphere is the sun), and so the mental file corresponding to the model is in fact also the mental file of a *Z*-representation. Then, depending on how one wants to speak, we can either say that *MS* exemplifies properties $P_1, \ldots P_\mathrm{n}$ (the properties that are highlighted in a given context of investigation), or we can say that C_M highlights the properties $P_1, \ldots, P_\mathrm{n}$. These properties can then be keyed up with properties $Q_1, \ldots, Q_\mathrm{m}$ to be imputed onto the target system. The elements of this account are summarised in Fig. 10.4.

10.7 Conclusions

We argued that anyone who subscribes to the idea that at least some models are non-concrete objects which nevertheless denote their targets faces the problem that denotation is a two-place relation that holds between existing objects. The identity

[14]It's worth noting here that we take the model's *representational* content to be the result of imputing the $Q_1, \ldots, Q_\mathrm{m}$ to *T*. this representational content is distinct from the model-content itself in so far as the latter concerns claims about target systems whereas the former concern claims about, or pretend about, the model systems themselves.

problem was to specify *what it is* that denotes the target system in cases of model-based representation. There are various options available here: one could adopt a realist account of what models are, and then attempt to accommodate those objects as the sorts of things which can denote. Alternatively, as we have done, one could adopt an antirealist position towards models. In the latter case, there are two further options available. One could attempt to deflate the notion of denotation in operation in scientific representation and adopt the view that scientific representation involves pretend-denotation (rather than denotation), or, alternatively, one could identify the objects of denotation elsewhere. The focus in this paper was on the latter approach. We have argued that we can identify models with their descriptions and their content generated in a Waltonian game of make belief. Our approach demonstrates that there is a coherent way of meeting the identification problem, by invoking model descriptions and their content, which captures much of what is valuable about indirect accounts of scientific representation like DEKI, and we hope to have at least indicated how other coherent answers could be developed.

There is a final thing to note. By identifying models with their descriptions plus the description's content it might seem like we have introduced an asymmetry into DEKI, and indeed many accounts of representation, with respect to how concrete and fictional models represent. In the case of concrete models, the concrete objects are themselves the objects of denotation: the wooden model-ship represents the real ship and the Phillips-Newlyn machine represents the economy (Frigg and Nguyen 2018). If we were to carry over the account presented in Sects. 10.4 and 10.6 to the context of concrete models (developing the picture displayed in Fig. 10.1), then symmetry with respect to how concrete and fictional models represent would seem to imply that it should be the descriptions of the concrete object which denote the target system, rather than the concrete object itself. This doesn't seem right. It would seem that it's the concrete object that is the model in those cases, not the descriptions thereof. If one were to insist on this, then DEKI (and with it other accounts of representation) would be asymmetric across the fictional/concrete model divide. If one were to accept it is model descriptions rather than objects that denote then one would have to reconsider how one thinks about concrete models.

One way of addressing this issue is to say that in the concrete case the model is a triple consisting of D, C, and the material object O. So, the model constructed by Philips and Newlyn was actually the machine plus a description of the machine along with the description's content. This suggestion is more natural than it might appear at first blush. In fact, it has been implicit in DEKI all along. DEKI says that a model is a model object under a certain interpretation (Frigg and Nguyen 2018, 213). Phillips and Newlyn's pipe system becomes a model when the flow of water is interpreted as the flow of money. This involves two steps. First one has to identify certain properties of the material object as relevant: the Phillips-Newlyn machine has to be construed as water-pipe-system rather than, say, as a plastic-and-metal system or a post-war-production system to become an economy-representation. In a second step the so-identified properties have to be connected to other properties to form an interpretation (in DEKI's sense), for instance by connecting the amount of water to the amount of money. None of this is in the material object *itself*. The scientists using the machine use a description to describe the machine as a water-pipe-system and then another description to interpret water properties in terms of economy properties. So, descriptions (and their content) really are part of the model, and including them in a material model is not just an ad hoc move to remove a tension between different parts of our account. So concrete models work like fictional models, just with a material object added to the unit we call "model".[15]

Finally, and possibly most importantly, whilst the identification problem has been solved, it remains to answer the relation problem. That models, identified with their descriptions and content, *can* denote target systems does not tell us *how* they do so. We have argued that our approach does not commit us to a descriptivist answer to this problem, but the answer itself remains to be explored.

References

Bach, K. (1987). *Thought and reference*. Oxford: Clarendon Press.

Baillargeon, R. (1995). Physical reasoning in infancy. In M. Gazzaniga (Ed.), *The cognitive neurosciences*. Cambridge, MA: MIT Press.

Barberousse, A., & Ludwig, P. (2009). Models as fictions. In M. Suárez (Ed.), *Fictions in science: Philosophical essays on modeling and idealization* (pp. 56–73). New York: Routledge.

Barrett, T. W., & Halvorson, H. (2016). Glymour and quine on theoretical equivalence. *Journal of Philosophical Logic, 45*(5), 467–483.

Carey, S. (2009). *The origion of concepts*. New York: Oxford University Press.

Contessa, G. (2007). Scientific representation, interpretation, and surrogative reasoning. *Philosophy of Science, 74*(1), 48–68.

Elgin, C. Z. (2010). Telling instances. In R. Frigg & M. C. Hunter (Eds.), *Beyond mimesis and convention: Representation in art and science* (pp. 1–18). Berlin/New York: Springer.

Elgin, C. Z. (2017). *True enough*. Cambridge, MA: MIT Press.

Fodor, J., & Pylyshyn, Z. (2014). *Minds without meanings: An essay on the content of concepts*. Cambridge, MA: MIT Press.

French, S., & Ladyman, J. (1999). Reinflating the semantic approach. *International Studies in the Philosophy of Science, 13*, 103–121.

Friend, S. (2011). The great beetle debate. *Philosophical Studies, 153*(183–211), 183–211.

Friend, S. (2014). Notions of nothing. In M. García-Carpintero & G. Martí (Eds.), *Empty representations: Reference and non-existence* (pp. 307–332). Oxford: Oxford University Press.

Frigg, R. (2010). Models and fiction. *Synthese, 172*(2), 251–268.

Frigg, R., & Nguyen, J. (2016). The fiction view of models reloaded. *The Monist, 99*, 225–242.

Frigg, R., & Nguyen, J. (2017). Models and representation. In L. Magnani & T. Bertolotti (Eds.), *Springer handbook of model-based science* (pp. 49–102). Cham: Springer.

Frigg, R., & Nguyen, J. (2018). The turn of the valve: Representing with material models. *European Journal for Philosophy of Science, 8*(2), 205–224.

Giere, R. N. (2010). An agent-based conception of models and scientific representation. *Synthese, 172*(1), 269–281.

[15] A difference is that *G* becomes obsolete in concrete models because the material object itself generates the model truths. When using the Phillips-Newlyn machine, we don't use principles of generation to find out how the economy behaves; we let the machine run!

Glymour, C. (1971). Theoretical realism and theoretical equivalence. In R. C. Buck, R. Carnap, & R. Cohen (Eds.), *Proceeding of the 1970 biennial meeting philosophy of science association* (pp. 275–288). Dordrecht: Reidel.
Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy, 21*(5), 725–740.
Grice, P. (1969). Vacuous names. In D. Davidson & J. Hintikka (Eds.), *Words and objections* (pp. 118–145). Dordrecht: Reidel.
Halvorson, H. (2012). What scientific theories could not be. *Philosophy of Science, 79*(2), 183–206.
Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. Amherst: University of Massachusetts.
Hendry, R. F., & Psillos, S. (2007). How to do things with theories: An interactive view of language and models in science. In J. Brzeziński, A. Klawiter, T. A. F. Kuipers, K. Łastowski, K. Paprzycka, & P. Przybysz (Eds.), *The courage of doing philosophy: Essays dedicated to Leszek Nowak* (pp. 123–157). New York: Rodopi.
Jeshion, R. (2010). Singular thought: Acquaintance, semantic instrumentalism and cognitivism. In R. Jeshion (Ed.), *New essays on singular thought* (pp. 105–140). Oxford: Calarendon Press.
Kamp, H., & Reyle, U. (1993). *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory* (Studies in linguistics and philosophy) (Vol. 42). Dordrecht: Springer.
Kaplan, D. (1990). Words. *Proceedings of the Aristotelian Society, Supplementary Volumes, 1*(64), 93–119.
Karttunen, L. (1976). Discourse referents. In J. McCawley (Ed.), *Notes from the linguistic underground* (Syntax and semantics) (Vol. 7, pp. 363–385).
Levy, A. (2015). Modeling without models. *Philosophical Studies, 152*(3), 781–798.
Murez, M., & Recanati, F. (2016). Mental files: An introduction. *Review of Philosophy and Psychology, 7*, 265–281.
Opie, I., & Opie, P. (1969). *Children's games in street and playground*. Oxford: Clarendon Press.
Perner, J., Huemer, M., & Leahy, B. (2015). Mental files and belief: A cognitive theory of how children represent belief and its intensionality. *Cognition, 145*, 77–88.
Perry, J. (1980). A problem about continued belief. *Pacific Philosophical Quarterly, 61*, 317–332.
Perry, J. (2001). *Reference and reflexivity*. Stanford: CSLI Publications.
Pylyshyn, Z. (2000). Situating vision in the world. *Trends in Cognitive Science, 4*, 197–207.
Pylyshyn, Z. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80*, 127–158.
Pylyshyn, Z. (2007). *Things and places: How the mind connects with the world*. Cambridge, MA: MIT Press/Bradford Books.
Quine, W. V. O. (1975). On empirically equivalent systems of the world. *Erkenntnis, 9*, 313–328.
Recanati, F. (2012). *Mental files*. Oxford: Oxford University Press.
Salis, F. (2013). Fictional names and the problem of intersubjective identification. *Dialectica, 67*(3), 283–301.
Salis, F. (2020). Capturing the Scientific Imagination, in P. Godfrey-Smith & A. Levy (Eds.), *The Scientific Imagination* (pp. 17–50). New York: Oxford University Press.
Salis, F., & Frigg, R. (2020). Capturing the scientific imagination. In P. Godfrey-Smith & A. Levy (Eds.), *The scientific imagination* (pp. 17–50). New York: Oxford University Press.
Slutkin, G. (2016). Is violence 'senseless'? Not according to science. Let's make sense of it and treat it like a disease. *Health Progress, 97*(4), 5–8.
Spelke, E. (1990). Principles of object perception. *Cognitive Science, 14*, 29–56.
Suppe, F. (2000). Theory identity. In W. H. Newton-Smith (Ed.), *A companion to the philosophy of science* (pp. 525–527). Oxford: Wiley-Blackwell.
Thomson-Jones, M. (2010). Missing systems and face value practise. *Synthese, 172*(2), 283–299.
Thomson-Jones, M. (2012). Modeling without mathematics. *Philosophy of Science, 79*(5), 761–772.

Toon, A. (2012). *Models as make-believe. Imagination, fiction and scientific representation*. Basingstoke: Palgrave Macmillan.

Walton, K. L. (1990). *Mimesis as make-believe: On the foundations of the representational arts*. Cambridge, MA: Harvard University Press.

Weisberg, M. (2007). Who is a modeler? *The British Journal for the Philosophy of Science, 58*, 207–233.

Worrall, J. (1984). An unreal image. *British Journal for the Philosophy of Science, 34*(1), 65–80.

Part IV
Fictionalism or Realism in Philosophy of Language

Chapter 11
Fictional Co-identification: The Explanatory Lightweight of Realism

Manuel García-Carpintero

Abstract Claims of fictional co-identification are ordinary prima facie true claims such as 'Ulysses is the same character as Odysseus'; they create the problem of explaining their truth-conditions. In previous work, I have defended a version of Yablo's figuralism – a purported irrealist view, on which reference to fictional characters is just hypostatization – a figure of speech. The irrealist credentials of the view could be questioned, however, because the metaphors it posits are pretty much "dead". And it might be thought that this is a good thing, because a realist view comes handy, for instance when it comes to account for fictional co-identification. In this paper I will show that this is illusory: ultimately, realist accounts of fictional co-identification must be grounded on irrealist pretense-theoretic views. This can be taken as evidence either for a "moderate realist" view along the creationist proposals of Thomasson, Voltolini, Walters, and others, or for a more straightforward vindication of irrealism.

Keywords Reference · Fictional reference · Intentional identity · Pretense

11.1 Introduction

Stacie Friend (2011, 2014) points out that we have intuitions of "co-identification" about, say, a debate confronting Nabokov, who asserts (2), with other critics, who had stated instead (1) – the novel just says that Samsa is transformed into a gigantic "vermin":

Financial support for my work was provided by the DGI, Spanish Government, research project FFI2016-80588-R, the award *ICREA Academia* for excellence in research, 2018, funded by the Generalitat de Catalunya, and from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement no. 675415, *Diaphora*. Thanks to Michael Maudsley for the grammatical revision.

M. García-Carpintero (✉)
Departament de Logos-BIAP, Department of Philosophy, University of Barcelona, Barcelona, Spain

J. L. Falguera, C. Martínez-Vidal (eds.), *Abstract Objects*, Synthese Library 422,
https://doi.org/10.1007/978-3-030-38242-1_11

(1) Critic: Gregor Samsa is transformed into a cockroach.
(2) Nabokov: No, Samsa is transformed into a beetle.

As Friend (2014, 308) indicates, Nabokov naturally "takes himself to identify the same character that Kafka invented and that his opponents misconstrue". How should we account for such intuitions, if we don't dismiss them as confused? Now, there is a natural connection between Friend's worry and Geach's problem of intentional identity: I take Friend's impressions of co-identification to constitute a particular case of Geach's (1967, 147) attitudes "with a common focus, whether or not there is something at that focus".[1] For we can adequately report Friend's case with a claim corresponding to Geach's famous "Hob-Nob" example, discussed in more detail in the next section:

(3) A critic thinks Samsa/a character of Kafka's was transformed into a cockroach, but Nabokov believes he/the same character was transformed into a beetle.

In her work, Friend provides an allegedly irrealist account of the phenomenon. Everett (2013) offers an alternative pretense-theoretic account. I myself favor an alternative account, which I also take to be irrealist, to be introduced below. The *prima facie* most natural account, however, appears to be realist; Salmon (2002, 2015) has advocated one along the realist lines he favors.

Underlying this discussion are issues concerning the nature of reference. In previous work, I have defended a "reference-fixing" form of descriptivism for expressions such as names and indexicals. Now, the very example that Friend uses is intended to put pressure on more straightforward forms of descriptivism, which she rejects, in favor of orthodox Millian views. The debate between Nabokov and the other critic shows that they associate different descriptive information with 'Samsa'; simple-minded descriptivist accounts would then wrongly imply that they are talking past each other. My ultimate goal is to show that the form of descriptivism that I hold doesn't succumb to this objection.[2] Here, however, this will mostly remain in the background; I'll only come back to the issue in my concluding remarks. I will limit myself to argue that any acceptable account of co-identification, realist or otherwise, will have to rely on an account entirely compatible with the sort of descriptivist view that my proposal envisages.

11.2 Intentional Identity

In setting up the problem posed by what he (Geach 1967) called *intentional identity*, Geach was interested in attitudes correctly ascribed by (4), on a reading that would be best captured in First Order Logic by (5) or (6) if we disown the ontological commitment associated with '∃' since Quine (1948):

[1] Pautz (2008, 149) also points this out. Cf. also the editors "Introduction" to Brock and Everett (2015), p. 6.

[2] García-Carpintero (2018) provides a recent presentation and references to previous work.

(4) Hob believes a witch has blighted Bob's mare, and Nob believes that she (the same witch) killed Cob's sow.
(5) $\exists\alpha$ (W(α) $\wedge$ Hob believes that B(α, b) $\wedge$ Nob believes that K(α, c)).
(6) $\exists\alpha$ (Hob believes that (W(α) $\wedge$ B(α, b)) $\wedge$ Nob believes that K(α, c)).

The interpretation that Geach pointed out is one on which (4) would be true in a world without witches, in which Hob and Nob are nonetheless thinking of the same non-existent witch, even though they do not know of each other's beliefs. The following scenario – taken from Glick (2012, 387), cf. also Edelberg (1986, 2) – provides intuitive grounds for such an interpretation: Hob and Nob live on opposite sides of town, and their social circles do not overlap at all; they have never heard of or encountered each other. But Hob and Nob have independently read and been persuaded by the local newspaper's claim that "Samantha, a witch, has been terrorizing the town." Each thinks he has discovered the cause of local livestock trouble, though the newspaper story was actually a complete fabrication.

Over the years, several writers have expressed skepticism about the availability of this reading (see Braun (2012) for a recent endorsement) but their doubts look unfounded to me. If Hob and Nob form their independent beliefs about an actually existing corrupt politician rather than about a witch reported in the newspaper (say, that the politician bribed Bob's son and that he wooed Cob's daughter, respectively), or if we happen to inhabit a world in which there in fact are witches, there is no problem in accepting the *de re* reading formalized by either (5) or (6).[3]

In both empty and non-empty cases, what seems to ground the relevant reading of the reports is (helping ourselves to Grice's (1969) already well-established metaphor) that the "mental dossiers" in both subjects' attitudes would be about the same individual, were one to exist.[4] Elaborating on Kripke's notion of a *communication chain* by means of which he provided a non- descriptivist outline of an account of reference-fixing, Burge (1983, 91–8) develops a notion of "quasi-anaphoric links", and justifies the Geachian intuitions in terms of it. Intersubjectively, as in Kripke's picture, they are constituted by intentions to use referential devices in accordance with the meaning that corresponding expressions have in the usage of the interlocutors on whom one relies.[5] Intrasubjectively, the relations between

[3] Braun (2012, 171–4) assumes that the analysis (5) is ontologically committal to witches, and uses their structural similarity to *de re* readings with actual witnesses to "explain" why the intuition that (4) is true on Geach's interpretation is confused. This is suspicious. Our intuitions about successful cases are reliable data for semantic theorizing, Braun assumes. Intuitions indistinguishable as such about referent-absent cases are instead confusions engendered by the similarity with the former, because he takes for granted that such readings are ontologically committal to witches. But methodologically it makes better sense to give the intuitions the same standing in all cases, explaining them by means of an account on which (5)/(6) are not so committal. This is what the version of irrealism I favor offers, as indicated below.

[4] Cf. Evans (1982, 362).

[5] As Sainsbury (2015, 205) puts it, "Acquisition involves a *de re* intention concerning that token. A self-conscious and sophisticated subject might inwardly think: that's something I will add to my repertoire, using it as those people do." He goes on to helpfully elaborate on what is involved in the more usual case in which the relevant intentions are merely implicit. Note, however, that for reasons

referential vehicles that mental files are intended to capture constitute such links.[6] Unlike Kripke's story, the links might involve referential expressions other than proper names; and they can ultimately depend on an unsuccessful act of reference.[7]

Like Burge's, most recent discussions dismiss skepticism, and attempt to develop the idea that (5) or (6) provides a good regimentation of the relevant truth-conditions, understood in such a way that it does not commit us to the actual existence of witches. Most recent proposals are consistent with Burge's appeal to quasi-anaphoric links, and provide elaborations of it.[8] Some adopt variants of what Edelberg (2006) calls *realism*: Edelberg (1986, 1992) and Cumming (2014) give accounts in terms of an ontology of "thought-objects", individuated by entities equivalent to Fregean senses, which might or might not coincide with actual referents; Salmon (2002) posits actually existing abstract objects; Glick (2012) and Pagin (2014) give modal accounts, taking the quantifier to range over possible objects including non-actual ones. Other approaches such as Priest's (2005) and Azzouni's (2010) question instead that quantifiers have the standardly assumed ontological commitments. Let's place these views in a wider setting.

11.3 Fictional Discourses: Realism and Irrealism

It will be convenient to distinguish three types of fictional discourse, which pose their own specific problems. Consider these sentences:

(7) When Gregor Samsa woke, he found himself transformed into a gigantic vermin.
(8) According to *Metamorphosis*, when Gregor Samsa woke, he found himself transformed into a gigantic vermin.
(9) Gregor Samsa is a fictional character.

Take firstly an utterance of (7) by Kafka, as part of the longer utterance by him of the full discourse which, with a measure of idealization, we can think constitutes the act of putting forward his creation *Metamorphosis* for us to enjoy.

I have elaborated elsewhere (García-Carpintero 2018), I dispense with his distinction between a syntactic (*using the same word*) and a semantic (*using the word in the same way*) components of the intention.

[6]In Sainsbury's helpful terms (*ibid.*), fulfilling the intention to use the same word "will result in acquiring a mechanism of name-reproduction, a mechanism that produces copies of the source name". He goes on to elaborate on the "mechanism" metaphor.

[7]The links are thus the relations of *de jure* or *internal coreference* that have recently received extensive discussion; cf. Fine (2007, 40, 68) (who speaks instead of objects "being represented as the same"), Lawlor (2010), Schroeter (2012), and references there.

[8]We should distinguish the debate about the semantics of ascriptions such as (4), from the one about the nature of the relations between the ascribed attitudes, even though they are related. For most writers, including Geach, the 'intentional identity' label covers both.

It is distinctive of such uses, which I will be calling *textual*,[9] that they are not intuitively truth-evaluable. The other two types differ in that they intuitively appear to be truth-evaluable. There is, firstly, the use of sentences such as (7) that we make when we are reporting the content of a fiction. I will call these content-reporting uses *paratextual*; according to Lewis (1978) and others, they are simply elliptic for intuitively equivalent ascriptions of propositional content such as (8). Finally, I will call the uses of sentences such as (9) *metatextual*; they are intuitively truth-evaluable but not directly content- reporting, in that they are not (or at least not obviously) equivalent to propositional content ascriptions like (8).

Kripke (2013) argues that a proper account of metatextual uses requires interpreting names such as 'Gregor Samsa' in them as referring to fictional entities. Van Inwagen (1977) provides an influential argument for such realism about fictional entities: a Quinean appeal to non- eliminable quantification over, and reference to, such entities in *prima facie* serious, truth- evaluable discourse, such as utterances of (9) and related metatextual uses in contexts of literary criticism. Such *ficta* could then be taken to be (exactly as in the options considered above for the realist account of the relevant readings of (3) and (4)) Meinongian non-existent entities, concrete non-actual *possibilia*, or (as both Kripke and van Inwagen prefer) abstract existent entities of various sorts, fully-fledged Platonic *abstracta* as in Wolterstorff (1980) or rather created artefacts, as in Salmon (1998), Thomasson (1999) and Schiffer (2003). Fictional entities of any of these sorts could also be invoked to account for either of the other uses, textual and paratextual, but this requires extra work; for neither of those entities can be straightforwardly taken to be the sort of thing capable of sleeping and waking.[10]

The intuitive obviousness of negative existentials involving fictional names ('Samsa doesn't exist') counts against non-Meinongian realist views, a point that Everett (2007, 2013, ch. 7) forcefully presses. He (2013, ch. 8) provides an interesting elaboration on equally well- known indeterminacy concerns about fictional realism, echoing Quine's (1948, 23) indictment: "the possible fat man in that doorway; and, again, the possible bald man in that doorway. Are they the same possible man, or two possible men? How do we decide? How many possible men are there in that doorway? Are there more possible thin ones than fat ones? How many of them are alike?" Everett (2013, ch. 7) and Sainsbury (2010, ch. 3 & 4) also articulate related problems for the Meinongian and possibilist alternatives.

As said, focusing on metatextual uses leads us naturally to think of the referential expressions in (7)–(9) as in fact referring when they are taken to make assertions. This might address qualms that Millians (those who take the referent of a name to exhaust is semantic content) would otherwise have to endorse the intuitive view that paratextual uses of (7) indeed make assertions – perhaps the ones explicitly made

[9] I take this and the other two related labels from Bonomi (2008).

[10] Voltolini (2006) provides a helpful exploration of the alternatives.

with (8). We could even entertain the view that textual uses of (7) make assertions too.[11] Focusing instead on textual uses leads to a contrasting irrealist picture.

When the creator of a work of fiction uses declarative sentences such as (7), or when she uses sentences of other types, we do not intuitively think of her as really performing the speech acts one typically performs with them in default contexts. In such cases, the sentences are used in some form of *pretense*, like the acts that actors perform on stage: they do not need to be drinking whisky, for they are merely pretending to do so; hence, we do not evaluate them by invoking any norms we would apply to non-pretend uses.

Perhaps the intuitively best option would be to combine fictional realism for the latter with a pretense-theoretic account of authors' uses of sentences like (7); this is in fact Kripke's (2013) "pluralist" suggestion, on which fictional names such as 'Gregor Samsa' have an empty, pretend use in (7), but a non-empty serious one in (9). In addition to the resulting profligacy, however, paratextual uses of (7) occupy a problematic middle ground for this ecumenical rapprochement. Also, as Everett (2013, 163–178) emphasizes, there are many mixed cases such as (10) below; for note that here whatever 'Gregor Samsa' designates is ascribed properties both from the internal, conniving paratextual perspective, but also from an external, metatextual viewpoint:

(10) At the start of *Metamorphosis*, Gregor Samsa – an emotional *alter ego* of himself vis-à- vis his father created by Kafka for his most popular novel – founds himself transformed into a gigantic vermin.

Everett takes these considerations as a good reason to extend the pretense-theoretic treatment to paratextual and metatextual uses. This, however, doesn't provide an immediately obvious account of negative existentials such as 'Samsa doesn't exist'; and there remains the intuitively strong impression that (7) in paratextual uses, (8) and (9) – together with (3) and (4) – make straightforward, truth-evaluable assertions.

Walters (ms) provides a compelling defense of Kripkean pluralism for names, combined with an artefactualist view of the referents of some such names, drawing on ideas also nicely articulated by Everett and Schroeder (2015). Walters rejects Millianism, assuming that empty names are nonetheless meaningful, and he then

[11] A view like this appears to follow from Ludlow's (2006) main claim, that in fictional contexts predicates such as 'is a vampire' (metonymically?) acquire an extended sense in which they truly, literally apply to the props representing vampires in the relevant fiction, such as actors playing vampire roles in *Buffy The Vampire Slayer*. Ludlow is not fully clear regarding what the props are in literary cases, but if we take them to be the representations to which fictional names refer in metafictional discourse on Walters' realist view presented below, the resulting proposal is a natural extension to textual discourse. (Walters himself nonetheless dismisses it, on the grounds I take it that the pretense view better accounts for our intuitions.) Martinich and Stroll (2007) defend a related view for textual uses, including those of sentences with apparently empty names like (7) – which, unlike in the previous proposal, they take to be in fact empty, without that preventing them from being true. They advance a performative view of the acts of the fiction-maker, which create "institutional facts" making their claims true.

extends a Waltonian, pretense-theoretic account of textual uses of (7) to paratextual uses. Against Walton (1990) and Everett (2013), however, Walters takes the likes of (8) to make truth-evaluable assertions, in which the use of the names is still the empty one; he assumes some non-Millian semantic account of propositional attitude ascriptions for that, although he grants to pretense theorists that it is the paratextual pretend use of (7) that grounds assertions such as (8), including empty names. In metatextual uses, however, we find according to him a non-empty homonym of the empty name that occurs in those other uses. It refers to a representation: intuitively, the (created) representation(−type) of Samsa which is a part of the whole representation of the fictional events portrayed in Kafka's *Metamorphosis*.[12] Walters then goes on to explain mixed cases like (10) on the assumption that they involve a form of independently well-attested metonymy-induced polysemy, as when we straightforwardly apply 'lion' to a representation of what literally, primarily is not a lion, like a statue of one; for we also naturally find similarly mixed cases here. Thus, a sculptor can say this of one of her creations:

(11) That lion is the best sculpture I've made this month; it is as ferocious as the one we saw yesterday at the zoo.

In previous work I have defended a similar package of views, but assuming a slightly different philosophical ideology. Like Walters, I argued that no adequate pretense-theoretic account can be happily combined with Millian views of singular reference, as in Walton's or Everett's proposals. This is not just for the reasons suggested by Walters; more fundamentally, we need to explain how the semantic content of (7) contributes to determine the content the fiction-maker proposes readers to imagine, or make-believe (García-Carpintero 2010a, 286–7). By relying on my own version of a non-Millian, descriptivism-friendly view of names and other referential expressions,[13] I have defended what I consider a form of irrealism for metatextual discourse: a version of Yablo's (2001) *figuralist* brand of fictionalism. on which the semantic referential apparatus (*de jure* directly referential expressions such as names and indexicals, quantifiers generalizing over the positions they occupy, expressions for identity) is used metaphorically in the likes of (3), deploying the figure of speech called *hypostatization* (García-Carpintero 2010b). It is a rather dead, conventionalized kind of metaphor, so, in contrast with pretense-theoretic fictionalist proposals, on this view utterances in metatextual discourse are straightforward assertions with truth-conditions.[14]

This might suggest that the view is after all realist, committed to referents of some sort for singular terms in metatextual discourse. I do not think so. One could follow Brock (2002), and claim that the literal content apparently involving commitment to fictional entities is in fact along the lines of (8): one about what

[12] Everett and Schroeder (2015) call such representations *ideas*.

[13] I refer again la reader to García-Carpintero (2018), which provides a recent presentation and further references.

[14] The pretense involved is hence *semantic* – as opposed to *pragmatic* – on Armour-Garb's and Woodbridge's (Armour-Garb and Woodbridge 2015) classification, if I understand it correctly.

is true according to a pretense – the pretense that some realist theory is true. Or – like Yablo himself – one could follow Walton (1993) in thinking that this applies in general to metaphors, which are a "prop-oriented" form of make-belief put forward with the aim of asserting a metaphorical content non-committal to fictional entities, through the process that Richard (2000) calls "piggybacking".[15]

My own preferred line, however, follows Yablo's (2014a) recent development of his views,[16] articulating the view that the truth of metatextual sentences including fictional names and their generalizations do not really commit us to the existence of fictional characters; for this is merely pretend-presupposed and, when we look at what they are really *about* (and hence the truth- makers for the claims we make with them) we do not find the referents they appear to pick out.[17] We find instead the "ideas for fictional characters" of Everett and Schroeder (2015), or Walters' (ms) representations thereof. So, as said, the differences between my preferred story and Walters' artefactualist view are perhaps rather insubstantial. We all end up interpreting (3), (4), (8) and (9) as making genuine assertions, whose truth is grounded on the pretenses thereof in textual and paratextual uses of (7).[18]

This is not to say that there are not substantive differences between the "easy ontology" view which I take lies in general behind Walters' realist proposal,[19] and the form of fictionalism I advocate. The recent exchange confronting Thomasson (2013), and Yablo (2014b) witnesses such differences, in spite of their both occupying intermediate position between heavy-duty Platonism and straightforwad eliminativism. On Thomasson approach (Thomasson 2015, 261), an uncontroversial claim such as (12) analytically entails (13), and hence (14), given "linking principles" which provide meaning-constitutive application conditions for the sortal *fictional character*:

(12) Kafka wrote a novel using the name 'Gregor Samsa' to pretend to refer to and describe a man.
(13) Kafka created a fictional character.
(14) There are fictional characters.

Given the analytical character of the entailment, Thomasson (2017, 775) contends that the inference from (12) to (13) is not really ampliative: although (13) "does involve us explicitly quantifying over entities that [(12)] doesn't mention", it "doesn't contain any 'new information'" not already contained in (12). However,

[15]This is what happens when the mother tells her child "the cowboy should now wash his hands for dinner"; i.e., it is to make an utterance which would be true-in-the-pretense if certain conditions obtained (mother and child are playing a game of cowboys and Indians, with specific principles of generation), with the intention of asserting such conditions (i.e., that the boy dressed as a cowboy now has certain obligations). Cf. Evans (1982, 363–4).

[16]Hoek (2018) provides a nicely precise, tight variation on these ideas.

[17]Cf. Cameron (2012) for a related view.

[18]Hoek (2018, §4) discusses in detail related cases, including the Geach sentences.

[19]Cf. Thomasson (2015) for a recent statement applied to the present case of fictional characters, and references there to earlier proponents.

as Yablo (2014b) points out, there are similar “linking principles” taking us, say, from claims about observable facts, to claims about theoretical entities explaining them; or from claims about how things perceptually look, to claims about how they are. Moreover, whatever good reason there might be to count Thomasson’s linking principles as features of the conceptual role of sortals such as ‘fictional character’, there are indiscernible reasons to count the linking principles in the alternative cases as aspects of the conceptual roles of the relevant concepts of theoretical entities or observable objects. All of this provides, I think, good reasons for preferring a fictionalist account, on which (14) is merely a pretend presupposition in (13), (13)’s literal content is not really asserted, and it is only something very much like (12) what the utterer of (13) is assertorically committed to.

Nevertheless, whatever the resolution of these debates, for my purposes here I can afford to take an ecumenical attitude towards “easy ontology” realist approaches to fictional characters like Thomasson’s and Walters’. What matters is that, against eliminativism, (3), (9) and (13) are true; and, against heavy-duty Platonism, what their truth ultimately comes to is that of uncontroversial claims such as (12). As we have just seen, Thomasson explicitly agrees that the informational content of claims like (3), (9) and (13) is just that of claims such as (12). It is indifferent for my present purposes whether this is so because fictional names in metatextual discourse refer to fictional characters, which *are* “ideas” or “representations” thereof in fictions – as moderate realists of Thomasson’s and Walters’s persuasion contend; or, as I prefer, rather because, although this is not so, talk of fictional characters is really figurative and the contents we assertorically commit ourselves to in making claims ostensively about fictional characters are only really about such ideas or representations thereof.[20]

11.4 The Cart and the Horse

I will now conclude by explaining why I can afford to remain ecumenical between those two views on metatextual discourse. The reason is that, even if we embraced a straightforward form of fictional realism, we would have the very same explanatory commitments about metatextual discourse that fictionalists and moderate realists explicitly embrace. In particular, we would still have to provide an elucidation of the semantics of judgments of co-identification independent of the posited objects, and appealing instead to their representations in the fiction. (This would then allow me to argue that only a descriptivist-friendly account of referential expressions, such as the one I myself hold, can deliver the goods; but I will not go into this here.) I

[20]Everett (2013, 143) nicely puts the main thought here: “I do not mean to deny that in some cases the entities invoked by certain fictional realists, who then go on to identify these entities with fictional characters, genuinely exist. My complaint is simply that, in these cases, the relevant entities are not fictional characters; the identification made is wrong”.

will develop this point by discussing Salmon's (2002) account, but, as I'll show, it equally applies to possibilist, Meinongian, or Platonist forms of realism.

Salmon explains the Geachian reading of (4) (and would hence analogously explain the judgment expressed by (3)) by contending that the longest-scope quantifier binding the objects of the attitudes ranges over abstract entities or mythical or fictional objects. The following articulates the proposal:

(15) There is a mythical witch such that (i) Hob believes: she has blighted Bob's mare; and (ii) Nob believes: she killed Cob's sow.

A mythical witch is an abstract, existent object, of the sort realists like Kripke, van Inwagen and the other writers previously mentioned take fictional objects to be. Now, Braun (2012) sensibly objects that (15) does not provide a good analysis of (4), because they have different semantic contents: (15) makes crucial use of an expression, 'mythical witch', whose (theoretically stipulated) semantic content differs from that of any expression in (4). I think a reasonable response to this objection can be provided on behalf of the realist proposal, as I am about to show[21]; but the response makes it manifest why realism on its own will not supply us with the full account of judgments of co-identification we are after – i.e., of why, exactly as under irrealist proposals, we still need an elucidation of claims such as (3) independent of realist posits.

To develop the response, I'll rely on two observations. The first observation is something we have already touched upon in the previous section, because Walters' moderate realist account of fictional characters heavily relies on it[22]: the existence of the phenomenon that Nunberg (1995) calls "meaning-transfer", as when we count lion-statues as falling under the extension of 'lion' ("I'll meet you by the lion at the square"), which (11) above illustrates. These metonymical extensions have sometimes become conventionalized aspects of the lexical meanings of expressions, accounting for some cases of polysemy; take, for instance, the type/token ambiguity for 'novel' and related expressions. For our present purposes we need not worry whether the relevant meanings should count as "semantic" (a feature of the lexical meaning of the expression) or rather "pragmatic" (a creative meaning that speakers manage to endow their utterances with, trading on the rationality of their audiences). The second observation was also touched upon in Sect. 11.2 above, to wit, that the Geachian reading of (4) is on a par with those cases in which there really is a witness for the binding existential quantifier in (5): (4) would also be true in a possible world in which there is a real witch about whom Hob and Nob both have certain attitudes.

Putting these two observations together, the defender of the realist analysis can reply as follows to Braun's objection. When we have the intuition that (4) is true in the Geachian reading, we are taking for granted that 'witch' might have one of its metonymically extended senses; thus, for instance, it would be true if the reports

[21] It is in fact a particular elaboration of the "modified Salmonian pragmatic theory" which Braun discusses in section 6 of his paper.

[22] Ludlow's (2006) more radical view mentioned in fn. 11 above, extending also to textual discourse, is another case in point.

about "Samantha" in the story we envisaged are in fact about a real woman who, even though not (strictly speaking) a witch at all, *is thought* to be one, or perhaps *represents herself* as being one.[23] Analogously, *mutatis mutandis*, the real object with witch-related features sufficient for it to count as a witch in the metonymically extended sense might not be any concrete entity, but an abstract one created together with the mythical theory of witchcraft. This sort of case might be the one obtaining in most Geachian cases. The claim that (15) provides a good account of (4) is not the (obviously unwarranted) claim that they are synonymous; but rather that it makes explicit an acceptable metonymical extension for 'witch' in utterances of (4).[24]

As I pointed out in the previous section, when developed in this way, few differences remain between this realist proposal and the figuralist one.[25] But the elaboration shows that, without further work, the realist proposal does not offer a satisfactory account of co-identification. For the suggested truth-conditions for (4) explain co-identification in terms of attitudes about the same "witch-like" abstract object; but we obviously still need to know what this amounts to: what has to be the case, for a witch-like abstract object to count as a witch? That questions like this need answers is reinforced by the fact that (in the typical case) we are talking about the attitudes of people who have no articulated theory of mythical objects, hence no explicitly articulated attitudes about them, but who would instead take their attitudes to be about real, concrete individuals.[26]

This is in fact a particular case of a well-known problem that realists of all stripes should confront, and this is why – as contended at the beginning of this section – it extends to all of them. Some of them want to say that content-reporting, paratextual uses of (7) can be straightforwardly true, without the need to take it in such uses as short for (8) – which, for most of them, in any case raises the same concerns as (7) does if assumed to include a non-referring expression, irrespective of its occurring in a subordinate clause. In fact all of them would need to say something like this, to deal with mixed cases such as (10). This is achieved by taking 'Gregor Samsa' to refer to one of the realist's posits. But this creates a problem: how can such entities fall under the extension of the predicate? Neither abstract nor non-existent objects literally are true verminous insects. Realists deal with this problem by either distinguishing two types of properties, or two types of predication.[27] On the latter proposal, for instance, the realist would say that the subject-predicate combination in (7) does not mean that the referent of the subject-term truly instantiates the

[23] Both Everett and Schroeder (2015) and Walters (ms) make related points.

[24] Salmon (2015) provides something close to this reply to Braun's criticism.

[25] For technical reasons I didn't mentioned in the previous section, the figuralist might well welcome "objects" witnessing the existential quantifiers – to belong, say, in the "outer domains" of the positive free logic she might want to rely on. However, as indicated there, assertoric commitment remains confined only on the fictionalist view to there being in the relevant fiction character- representations that make true the relevant claims.

[26] Everett (2013, 179–188) develops a similar point.

[27] Cf. Everett (2013, 170–7) for a good discussion of the two options and their problems, and references to the original works articulating the strategies.

property expressed by the predicate, but merely, say, that such property *is ascribed to it* in some fiction. This raises the legitimate request to explain how fictions ascribe properties to the sort of object posited by realist theories, given that typically neither their creators nor their intended audiences have an elaborated view of them – and to do it without mentioning the posited objects themselves; and this is simply the challenge raised in the previous paragraph, in a more general form.

I think that the realist can also plausibly answer this more general form of the challenge; in fact, once more, as I'll presently grant, what she can say will be very close to what the figuralist approach affords. But the challenge must be met, which was the present point: just positing a Meinongian non-existent object, a *possibile*, or an abstract entity to witness the existential quantifier in (3), without further explanatory work, will not ease our qualms. We still need to explain what is it to have attitudes about a common witch-like Meinongian non-existent, *possible*, abstract entity, or in general about a single witch-in-an-extended-sense "entity"; and we need to explain this independently of these posits.[28]

Note that, as Everett and Schroeder (2015, 289–91) point out, irrealists about metatextual discourse are in the same boat, for variations on the questions raised here for realists can be legitimately posed to them. They ask (*ibid*., 290), "Why is it 'correct' to pretend that Holmes is a character created by Conan Doyle but not that Holmes is a character created by Tolstoy? Moreover, what is the sense of debating whether Holmes is a one-dimensional character if we are merely debating what shape our pretence or presupposition should take?" Obviously, the fictionalist will need to answer this question by pointing out that metatextual assertions made in pretense about Holmes involve assertoric commitments about what is going on in Conan Doyle stories in general, and the Holmes-representations to be found there in particular – exactly as (3), (9), and (10) are to be understood as involving assertoric commitments about Samsa- representations to be found in *Metamorphosis*.

This is a cart-before-the-horse concern for fictional realism. What is needed is an account of aboutness or reference that uniformly applies to straightforward assertions, such as those in metatextual claims, and the pretenses of textual discourses, on which our account of co-identification can then rely; i.e., an account of aboutness, reference and co-reference in textual discourse, i.e., in the paradigmatic cases that render themselves to the pretense-theoretic account. Afterwards we can debate whether judgments of co-identification involving fictional characters are exactly the same thing as those involving real particulars, as realist contend, or should be considered as involving some form of metaphor or pretense, as I prefer to say. Our target utterance (3) is a piece of metatextual discourse. Whether we give it a realist treatment like Salmon's (2002), or the figuralist one I favor, its truth-conditions are dependent on the aboutness of the original representations to which the relied-upon

[28] I do not mean to suggest that realists have not taken up this challenge; Thomasson (1999, 90–91) makes a proposal for her created-abstract-object form of realism compatible with my proposal below, something similar could be invoked on behalf of a Meinongian proposal, as elaborated by Priest (2011), and both Everett and Schroeder (2015) and Walters (ms) elaborate on how their brand of moderate realism deals with it.

"quasi-anaphoric links" ultimately lead – in our case, the relevant pieces of textual discourse.

11.5 Conclusion

In this paper I have taken up Stacie friend's challenge to descriptivist accounts of (fictional) names based on the fact that we make true claims of co-identification involving such names.

Although the account of such (metatextual) claims I have assumed has important similitudes with moderate realist proposals, its figurativist character makes my stance highly sympathetic to the irrealist assumptions about fictional characters that Friend invokes in setting up the problem. But I have here elaborated on a point that Evans (1982, 367) makes against fictional-character realism, namely, that a proper account of paratextual and metatextual discourses invoking them should nonetheless explain the connection between uses of names in such discourses and their uses in textual discourses and uses of names as genuine singular terms. Although for reasons of space I didn't elaborate on this here, if we went into the details concerning how aboutness in textual discouse grounds aboutness in metatextual speech acts, we would see that the form of descriptivism I myself have defended is not only compatible with Friend's points, but in fact offer the best account of them.

References

Armour-Garb, B., & Woodbridge, J. (2015). *Pretense and pathology*. Cambridge: CUP.

Azzouni, J. (2010). *Talking about nothing*. Oxford: Oxford University Press.

Bonomi, A. (2008). Fictional contexts. In P. Bouquet, L. Serafini, & R. Thomason (Eds.), *Perspectives on context* (pp. 213–248). Stanford: CSLI Publications.

Braun, D. (2012). Hob, nob and mythical witches. In W. P. Kabasenche, M. O'Rourke, & M. H. Slater (Eds.), *Reference and referring* (pp. 149–187). Cambridge, MA: MIT Press.

Brock, S. (2002). Fictionalism about fictional characters. *Noûs, 36*, 1–21.

Brock, S., & Everett, A. (2015). *Fictional objects*. Oxford: Oxford University Press.

Burge, T. (1983). Russell's problem and intentional identity. In J. E. Tomberlin (Ed.), *Agent, language, and the structure of the world. Essays presented to Hector-Neri Castañeda, with his replies* (pp. 79–110). Hackett: Indianapolis.

Cameron, R. (2012). How to be a nominalist and a fictional realist. In C. M. Uidhir (Ed.), *Art and abstract objects* (pp. 179–196). Oxford: Oxford University Press.

Cumming, S. (2014). Indefinites and intentional identity. *Philosophical Studies, 168*, 371–395. https://doi.org/10.1007/s11098-013-0131-9.

Edelberg, W. (1986). A new puzzle about intentional identity. *Journal of Philosophical Logic, 15*, 1–25.

Edelberg, W. (1992). Intentional identity and the attitudes. *Linguistics and Philosophy, 15*(6), 561–596.

Edelberg, W. (2006). Intersubjective intentional identity. *Journal of Philosophy, 103*, 481–502.

Evans, G. (1982). *The varieties of reference*. Oxford: Clarendon Press.

Everett, A. (2007). Pretense, existence, and fictional objects. *Philosophy and Phenomenological Research, 74*, 56–80.

Everett, A. (2013). *The nonexistent*. Oxford: OUP.

Everett, A., & Schroeder, T. (2015). Ideas for stories. In S. Brock & A. Everett (Eds.), *Fictional objects* (pp. 275–293). Oxford: Oxford University Press.

Fine, K. (2007). *Semantic relationalisms*. Oxford: Blackwell.

Friend, S. (2011). The great beetle debate: A study in imagining with proper names. *Philosophical Studies, 153*, 183–211.

Friend, S. (2014). Notions of nothing. In M. García-Carpintero & G. Martí (Eds.), *Empty representations: Reference and non-existence* (pp. 307–332). Oxford: Oxford University Press.

García-Carpintero, M. (2010a). Fictional singular imaginings. In R. Jeshion (Ed.), *New essays on singular thought* (pp. 273–299). Oxford: Oxford University Press.

García-Carpintero, M. (2010b). Fictional entities, theoretical models and figurative truth. In R. Frigg & M. Hunter (Eds.), *Beyond mimesis and convention – Representation in art and science* (pp. 139–168). Dordrecht: Springer.

García-Carpintero, M. (2018). The mill-Frege theory of proper names. *Mind* 127(508), 1107–1168. https://doi.org/10.1093/mind/fzx010.

Geach, P. (1967). Intentional identity. *Journal of Philosophy, 64*, 27–32. Also in *Logic Matters*, Blackwell, Oxford, 1971, 146–53.

Glick, E. (2012). A model approach to intentional identity. *Noûs, 46*(2012), 386–399.

Grice, H. P. (1969). Vacuous names. In D. Davidson & J. Hintikka (Eds.), *Words and objections* (pp. 118–145). Dordrecht: D. Reidel Publishing Company.

Hoek, D. (2018). Conversational exculpature. *Philosophical Review, 127*(2), 151–196.

Kripke, S. (2013). *Reference and existence*. Oxford: Oxford University Press.

Lawlor, K. (2010). Varieties of coreference. *Philosophy and Phenomenological Research, 81*, 485–495.

Lewis, D. (1978). Truth in fiction. *American Philosophical Quarterly, 15*, 37–46. Reprinted with postscripts in D. Lewis, Philosophical Papers, vol. 1, pp. 261–280, Oxford: Oxford University Press, 1983.

Ludlow, P. (2006). From Sherlock Holmes and Buffy to Klingon and Norrathian platinum pieces: Prfetence, contextualism, and the myth of fiction. *Philosophical Issues, 16*, 162–183.

Martinich, A. P., & Stroll, A. (2007). *Much ado about nonexistence: Fiction and reference*. Lanham: Rowman and Littlefield.

Nunberg, G. (1995). Transfers of meaning. *Journal of Semantics, 12*, 109–132.

Pagin, P. (2014). Intersubjective intentional identity. In M. García-Carpintero & G. Martí (Eds.), *Empty representations: Reference and non-existence* (pp. 91–113). Oxford: OUP.

Pautz, A. B. (2008). Fictional coreference as a problem for the pretense theory. *Philosophical Studies, 141*, 147–156.

Priest, G. (2005). *Towards Non-Being*. Oxford: Clarendon Press.

Priest, G. (2011). Creating non-existents. In F. Lihoreau (Ed.), *Truth in fiction* (pp. 107–118). Ontos Verlag: Frankfurt.

Quine, W. V. O. (1948). On what there is. *Review of Metaphysics, 2*, 21–38. Reprinted in in his From a Logical Point of View, 2nd edition, revised, Cambridge, MA: Harvard University Press, pp. 1–19.

Richard, M. (2000). Semantic pretence. In A. Everett & T. Hofweber (Eds.), *Empty names, fiction and the puzzles of non-existence* (pp. 205–232). Stanford: CSLI.

Sainsbury, M. (2010). *Fiction and fictionalism*. London: Routledge.

Sainsbury, M. (2015). The same name. *Erkenntnis, 80*, 195–214.

Salmon, N. (1998). Nonexistence. *Noûs, 32*, 277–319.

Salmon, N. (2002). Mythical objects. In J. K. Campbell, D. Shier, & M. O'Rourke (Eds.), *Meaning and truth: Investigations in philosophical semantics* (pp. 105–123). New York: Seven Bridges.

Salmon, N. (2015). The philosopher's stone and other mythical objects. In S. Brock & A. Everett (Eds.), *Fictional objects* (pp. 114–128). Oxford: Oxford University Press.

Schiffer, S. (2003). *The things we mean*. Oxford: Clarendon Press.

Schroeter, L. (2012). Bootstrapping our way to samesaying, *Synthese* Online First, https://doi.org/10.1007/s11229-012-0099-6.

Thomasson, A. (1999). *Fiction and metaphysics*. Cambridge: Cambridge U.P.

Thomasson, A. L. (2013). Fictionalism versus deflationism. *Mind, 122*, 1023–1051.

Thomasson, A. L. (2015). Fictional discourse and fictionalisms. In S. Brock & A. Everett (Eds.), *Fictional objects* (pp. 255–274). Oxford: Oxford University Press.

Thomasson, A. L. (2017). Why we should still take it easy. *Mind, 126*, 769–779.

van Inwagen, P. (1977). Creatures of fiction. *American Philosophical Quarterly, 14*, 299–308.

Voltolini, A. (2006). *How Ficta follow fiction. A syncretistic account of fictional entities*. Dordrecht: Springer.

Walters, L. (ms). *Fictional names*. http://www.leewaltersphilosophy.co.uk/research.html.

Walton, K. (1990). *Mimesis and make-believe*. Cambridge, MA: Harvard U.P.

Walton, K. (1993). Metaphor and prop oriented make-believe. *European Journal of Philosophy, 1*, 39–56.

Wolterstorff, N. (1980). *Works and worlds of art*. Oxford: Clarendon Press.

Yablo, S. (2001). Go figure: A path through fictionalism. In P. A. French & H. K. Wettstein (Eds.), *Midwest studies in philosophy XXV* (pp. 72–102). Oxford: Blackwell.

Yablo, S. (2014a). *Aboutness*. Princeton: Princeton University Press.

Yablo, S. (2014b). Carnap's paradox and easy ontology. *Journal of Philosophy CXI, 9*(10), 470–501.

Chapter 12
What Is the Difference Between Hamlet and Me? Fiction, Metaphysics and the Nature of Our Moral Thinking

Sofia Miguens

The key to seeing the centrality of fiction in metaphysics lies in (. . .) recognizing the similarities between fictional objects and other entities.

A. Thomasson, *Fiction and Metaphysics*

Abstract Starting from the main current views this essay considers whether entities (e.g. characters) in fiction should be viewed as abstract objects. I highlight some features of the historical concrete-abstract distinction, and, in particular, how entities in fiction are involved in our moral thinking. Here I call attention to an aspect of moral thinking orthogonal to that which currently divides moral realists and moral fictionalists and sketch an argument for fictional entities being in a specific sense concrete. Although the article does not, *per se*, call into question the approach to metaphysics according to which fictional characters are not like concrete objects, it exemplifies how a different perspective on the job of metaphysics (e.g. Cora Diamond's realistic spirit) gives the problem of fictional entities a radically different shape.

Keywords Fictional entities · Metaphysics · Concrete-abstract distinction · Moral thinking

12.1 Frameworks

Let us consider a fictional entity, say, Shakespeare's Hamlet. What, or who, is Hamlet, the troubled Prince of Denmark? Does he even *exist*? One reason in favour of answering this last question affirmatively is the fact that thousands of people in the last 400 years have read Shakespeare's play and have thought about Hamlet

S. Miguens (✉)
University of Porto, Porto, Portugal

J. L. Falguera, C. Martínez-Vidal (eds.), *Abstract Objects*, Synthese Library 422,
https://doi.org/10.1007/978-3-030-38242-1_12

in order to think about themselves, or the human condition. Yet is this sufficient for the claim that Hamlet exists? And what kind of question is this anyway? What leads us to pose it? In what follows I look into what I believe are some limits of standard current discussions of the nature of fictional entities taking another literary character, Elizabeth Costello, as my reference.

Elizabeth Costello is a character in works of South African Nobel Laureate J. M. Coetzee. She first appeared in his Tanner Lectures, which were given in 1997–98[1]. Coetzee's deliverance of the Tanner Lectures took the form of a story, a story where an aging Australian writer, Elizabeth Costello, is herself giving a set of lectures. The lectures (Coetzee's) and a set of comments were then published as *The Lives of Animals.* The book was regarded as a moral argument within a fictional framework. In the 2003 novel *Elizabeth Costello* Coetzee once again takes up the character of Costello. He goes on describing her. She is an elderly woman who is haunted by the horror of what we do to other animals. She is wounded by her knowledge of what goes on in the way humans treat other animals. She likens it to the Holocaust, in a way many people around her consider extremely offensive. She is also wounded by the knowledge of how unhaunted others are by this knowledge which haunts her. Elizabeth Costello is no militant vegetarian or moralist; her intention is not to persuade other people of her reasons through the best arguments she can find – all she wants is, she says, to save her soul. Also she is totally aware of the inconsistencies between her feelings and her behavior. In the words of American philosopher Cora Diamond,[2] Costello is *unplain*, not simple.[3] So we have a character, a fictional entity, to start with. I now want to ask: who, or what, is Elizabeth Costello, according to standard current approaches to the nature of fictional entities?

Possibilism, neo-Meinongianism and artefactualism are frameworks for answering the question. According to possibilism, fictional entities are *possibilia* – they are just like actual objects, except that they do not exist in the *actual world* but only in some other *possible worlds*. In the worlds in which they exist they are fully determined. This is so even if the stories, such as the stories where Elizabeth Costello appears as a character, do not themselves fill in all details about Elizabeth Costello. Possibilism supposes a commitment to realism about possible worlds, which should be justified independently. It also faces the problem that given the incompleteness of fictional entities in the stories where they appear there will be a multiplicity of Elizabeth Costello-candidates. I.e. there is more than one possible world in which J. M. Coetzee's stories involving Elizabeth Costello are fact, and in which there is a writer, an elderly woman haunted by the horror of what we do to animals and does the things recorded of her in Coetzee's stories; not all of

[1] Coetzee 1999.

[2] Diamond 2008.

[3] Diamond's term has an intentionally archaic flavour.

these Elizabeth Costello-candidates are the same. In fact David Lewis' view of counterparts, a view according to which specific individuals are world-bound, i.e. exist in only one world (in other possible worlds there are only counterparts), is one way of avoiding such problem, a problem which arises for S. Kripke's conception of possible worlds.[4]

Possibilism is one way to answer our question. Neo-Meinongianism is another. According to Neo-Meinongianism,[5] fictional entities are Meinongian objects. One gets to Meinongian objects from a condition: the condition that for every describable consistent set of properties, there is an object corresponding to such properties. Fictional entities are objects given in terms of the properties they have in the stories that feature them. Because Coetzee's writings are quiet on these matters, on the Meinongian model Elizabeth Costello is not right-handed; nor is she left-handed; nor is she ambidextrous. For orthodox neo-Meinongians (e.g. Terence Parsons and Dale Jacquette[6]), Elizabeth Costello is a concrete, albeit nonexistent, correlate of some specific set of properties: she is an aging writer, an Australian, and she has a son called John, who teaches physics and astronomy at a college in Massachussets, and so on, where the *and so on* includes all the properties *P* such that it is true in the Elizabeth Costello stories that Elizabeth Costello has *P* (remember that Coetzee himself writes more than just one story about her). While agreeing that fictional entities are a subset of Meinongian objects, *unorthodox* neo-Meinongians such as Edward Zalta (about whom I will say more in what follows) maintain that they should be conceived of as abstract rather than concrete. For any collection of properties there is indeed an individual that has all these properties. Still, an object may *exemplify* or not the properties that it *encodes*. Fictional entities are those abstract objects which *encode* properties while ordinary *individuals* (such as me and you) *exemplify* them. Concrete objects are identified and individuated in terms of their spatiotemporal location, whereas abstract objects, since they are not the kind of thing that could have a location in spacetime, must be identified and individuated in some other way. Unlike fictional entities conceived on the model of possibilism, fictional entities so conceived are not completely determined with respect to their properties: for Zalta they are something like generic objects or roles, along the model of Platonic attributes.

Yet another answer is artefactualism. According to artefactualism fictional entities are *abstract artifacts,* a hereto disregarded ontological category. This is the well known thesis of Amie Thomasson in her *Fiction and Metaphysics* (1999). They, as it were, come into being once they are conceived by their authors; to that extent, they are authorial creations. Like unorthodox neo-Meinongians, artefactualists believe

[4]Lewis 1986.

[5]In reference to Austrian philosopher Alexius Meinong (1853–1920), who was famous for his ontology. Meinong was a student of Franz Brentano (1838–1917).

[6]See e.g. Parsons 1980 and Jacquette 1996.

that fictional entities are abstract entities. Yet contrary to what Meinongians claim, fictional entities are created at a certain point in time; not merely discovered or picked out. Not only do they have a beginning in time, but also they are *existentially dependent* entities, as all cultural and institutional entities are. Roughly speaking, an entity O existentially depends on another entity O' if O couldn't exist without O' existing. More specifically, fictional entities depend historically rigidly on the authors who create them (necessarily, if O comes into being at t, then the author(s) who creates O exists at some time t' before t) and constantly depend generically on the literary works that feature them (necessarily, if O goes on existing, then some literary work or other featuring O exists during O's time of existence). While historical rigid dependence accounts for a fictional object's coming into being, constant generic dependence accounts for its continued existence or persistence (Thomasson 1999 includes an extended discussion of such dependencies).

As I said, possibilism, Meinongianism and artefactualism each outline an answer to the question regarding who, or what, Hamlet, or Elizabeth Costello, or other fictional entities, are. My intention in this article is not so much to engage in the ongoing discussion thus framed[7] but rather to consider some aspects and presuppositions of the discussion itself which I think are revealing of choices of how to do metaphysics.[8]

[7]If we chose to go this way, I believe Amie Thomasson is right in seeing Neo-Meinogianism as the main current alternative to artefactualism. For a current sophisticated combination of artefactualism and Neo-Meinongianism see Voltolini 2015. According to his syncretistic metaphysics *ficta* are hybrid entities individuated in terms of both a certain makebelieve narrative process and the set of properties that one such narration mobilizes.

[8]I am thankful to the anonymous reviewer who pointed out that nothing in the present article *per se* calls into question the approach to metaphysics according to which fictional characters are not like concrete objects. He/she is absolutely right: in order to do that I would have to go through Cora Diamond's criticism of the whole enterprise of analytic metaphysics from the viewpoint of the realistic spirit and I will not do it in what follows (I try to do it e.g. in Miguens 2019, *Uma leitura da filosofia contemporânea – Figuras e Movimentos*, Lisboa, Edições 70, pp. 374–385, and Miguens 2011, Miguens 2017). The ambition of this article is much more limited: here I am simply exploring the consequences of such criticism in a conception of fictional characters. What's more, this is not a central topic in Diamond. What is a central topic for her is the nature of moral thinking, and the importance of literature for ethical experience and imagination in the context of the realistic spirit. Given all these constraints, the article's main points could be considered somehow distant from the central topic of the volume. Still, I find it relevant to exemplify how from a perspective on metaphysics different from mainstream analytic metaphysics the topic of fictional entities radically changes shape.

The anonymous reviewer also suggests that that it would be sufficient to say that Hamlet, or any other fictional entity, is abstract in one sense, the sense that concerns the analytic metaphysician, and that Hamlet is not abstract in another sense, the sense that concerns Cora Diamond, or just use different terms (e.g. there's the abstract/concrete distinction, and there's the flat character/round character distinction). That is certainly one possible way to go but it would miss my main point in this article which is to show that the problem of fictional entities, once one changes perspective regarding the job of the metaphysician, is not a self-contained problem in metaphysics anymore but a question relevant to moral thinking.

12.2 A Challenge

Discussions regarding fictional entities often take place within general discussions regarding abstract objects, themselves taking place within metaphysical discussions concerning questions of existence and persistence. At the background of at least some such discussions is the need to account for practices, such as scientific practices, as they involve non-spatiotemporal entities, such as in mathematics. Some hold that mathematical entities are indispensable for scientific explanation. But entities in fiction, unlike mathematical entities, are not involved in scientific explanation,[9] so apparently there is no room for the indispensability arguments familiar in such contexts. I want to point out, though, that entities in fiction *are* indeed involved in our cognitive practices – only these are practices other than scientific practices, they are e.g. moral practices. As I said at the start, in the last 400 years thousands of people have thought about themselves, and others, or, as we might say, the human condition, by coming across Shakespeare's play and thinking about Hamlet. We think about ourselves (and about the ways we think about ourselves) in engaging with Elizabeth Costello. One might take it that understanding the nature of such involvement should be at stake in the discussion of fiction and not just the naked question of whether fictional entities exist or not, whether they are changing of unchanging, physical or non-physical, mental or non-mental.

With that thought in mind, I want to bring in a question which lies at the intersection of moral philosophy and the philosophy of literature, and which is thus different from the questions regarding the nature of fiction and of truth and reference and identification in fiction which usually go together with the discussion of the metaphysics of fictional entities.[10] Cora Diamond poses the question in her *The Difficulty of Reality*, taking as example *The Lives of Animals*. As I said before, Costello is the central character of Coetzee's contribution to that volume, which is followed by comments by several people, including Australian ethicist Peter Singer.

In *The Lives of Animals* Peter Singer, as other commentators of Coetzee's lectures, takes the question at stake to be how human beings should treat other animals. The character Elizabeth Costello is, as it were, a device for presenting an argument (in this case an argument in favour of radical egalitarianism), which may then be assessed as a good or bad argument:

> All beings with interests are deserving of moral regard
> Non-human animals are beings with interests
> Non-human animals are deserving of moral regard.

[9]Of course if one is a fictionalist regarding the nature of mathematical entities things go a very different way: a mathematical fictionalist such as Hartry Field, maintains that there are zero abstract entities. In such circumstances not even mathematical entities – the paradigm case in the discussion of abstract objects –, are taken to be existing abstract objects.

[10]For a recent and very thorough and useful discussion of such questions, see García-Carpintero 2016.

Diamond's question, the question I am interested in, is the following: *Are fictional characters simply devices for presenting arguments in the frame of fiction, or are they something else*?[11]

The way Singer sees things, the character Elizabeth Costello is simply an *Ersatz* for the argument above; it could be replaced by it at no cost. It is as such that literature may matter for moral philosophy according to Singer: characters and stories do no more than what could be done by argument. Diamond believes thinking as Singer does gets things exactly wrong about the nature of our ethical thinking – that is simply not how literature is involved in it. Elizabeth Costello is not an *Ersatz* for an argument, she is not a place-holder. She cannot easily be replaced by an argument; in fact she simply could not be replaced, one main reason being that she is *unplain*. In other words, because she is a character and not an argument, what she is, what she thinks, what she is proposing, is not even fully spelled out, and is certainly not without ambiguities and contradictions–; this is one reason why the idea that the argument that could replace the character is simplistic. What is actually happening is something else. By engaging with her we are changing our ways of seeing things – in this case we are putting to test what we think being an animal is. This naturally bears on the way we think of what being a human is. Thus we are putting to test the way we think about what kind of being we ourselves are. In more general terms, we are putting to test the resistance reality poses to our thinking one way or another about what being an animal and being a human animal is. In other words, we are not going through arguments about animal rights in the framework of fiction– things are not that simple. Arguments about animal rights may run only when we have already settled what being an animal is. The resistance reality poses to our way of conceiving things, which calls for an ongoing work of seeing things in new ways, is what Diamond calls *the difficulty of reality*. With Costello, Diamond says «Coetzee gives us a view of a profound disturbance of soul, and puts that view into a complex context. What is done by doing so he cannot tell us, he does not know. What response we may have to the difficulties of the lectures, to the difficulties of reality, the lectures themselves are not supposed to settle. This itself expresses a mode of understanding of the kind of animal we are, and indeed of the moral life of this kind of animal».[12]

So let us get back to fictional entities, keeping in mind the extra challenge of having a realistic conception of our moral thinking, namely when our moral thinking engages with fiction, and with unplain characters such as Elizabeth Costello, herself according to Diamond a 'wounded animal', immensely conscious of the limits of thinking and of argumentation.[13] In that, we might agree, she is more like me than, say, like the number 9.

[11]This is a confrontation I was interested in for other reasons (namely thinking about the nature of moral thinking). See Miguens 2011 and Miguens 2017.

[12]Diamond 2008: 56.

[13]Diamond 2008: 52.

12.3 A View of Metaphysics as a Theory of Objects. What Is Going on with ‘Concrete’ and ‘Abstract’

When we claim that entities in fiction are either concrete or abstract what reasons do we have to think one way or another? What, in fact, do we mean by *abstract* and *concrete*?

It does seem like a quite straightforward, even very attractive, thing to say about fictional entities (such as Hamlet, Sherlock Holmes or Elizabeth Costello) that they are abstract objects, in the sense of being *non-spatiotemporal*. In the terms of Edward Zalta’s Neo-Meinongian theory of abstract objects, fictional entities are those abstract objects which *encode* properties while ordinary *individuals* (such as me and you) simply *exemplify* them. Concrete objects are identified and individuated in terms of their spatiotemporal location, whereas abstract objects, since they are not the kind of thing that could have a location in spacetime, must be identified and individuated in some other way. Needless to say, the nature of particulars is as much a problem in metaphysics as the nature of universals. Particulars – such as this computer in front of me now, or myself, or this chair I am sitting on – may be taken to be *substracta*, or bundles, or substances – again, there must be independent reasons for thinking in one way or another. Anyway it is not a conception of particulars but rather a distinction between two fundamental kinds of predication that is essential to understand Zalta’s unorthodox Neo-Meinongian proposal about fictional entities.

The distinction is formally represented in the theory as the distinction between the atomic formulas ‘*Fx*’ (‘*x* exemplifies *F*’) and ‘*xF*’ (‘*x* encodes *F*’). The formula ‘*Fx*’ represents the classical kind of predication; it is used to logically analyse such simple sentences as ‘John is happy’, ‘Clinton is president’ or ‘Cora Diamond is a philosopher’. A student of Meinong, Ernest Mally (1879–1944), had already put forward the idea that sentences about fictional objects, such as ‘Sherlock Holmes is a detective’, ‘King Lear had 3 daughters’, or ‘EC is a writer’ should not be represented in terms of the notation ‘*Fx*’. Only real, concrete, objects can exemplify properties such as being a detective, having daughters or being a writer. Mally thought that there must be some mode of predication, some sense of the words ‘is’ and ‘has’, for which it is true to say ‘Sherlock Holmes is a detective’ and ‘Elizabeth Costello is a writer’ or ‘Lear has three daughters’; we wouldn’t understand stories properly, as we do, if we did not imagine objects that, in some sense, were instances of the properties in question. That is why Mally introduced the notion ‘*x encodes F*’ as a new mode of predication, a mode of predication appropriate for the logical analysis of sentences about entities in fiction and other abstract objects. Encoding thus provides the means for identifying and individuating abstract objects.

As I mentioned above, not all neo-Meinongians agree with Zalta in taking fictional entities to be abstract: C-Meinongians,[14] like Terence Parsons or Dale Jacquette, take fictional entities to be non-existing concrete entities. They are taken to be concrete in the sense that they are *correlates of the exact same properties*

[14] See Salis 2013.

which are predicated of common spatiotemporal objects. The discussion within neo-Meinongianism is the context of Zalta's approach to fictional entities. I will thus focus on it to bring forth some assumptions regarding 'concreteness' and 'abstractness'.

Zalta's theory of abstract objects is a metaphysical theory and in fact the expression of an engagement in a research programme in metaphysics. It is a metaphysical theory which conceives of itself as a theory of objects, a *Gegenstandtheorie,* such as that of Meinong himself. A *Gegenstandtheorie* aims at a classification of objects. In such theory fictional objects fall neatly in a classification of objects, which includes from numbers to propositions, *abstracta* and possible objects, future objects, even the actual world. Here is Zalta being explicit about his choice as how to do metaphysics:

> Metaphysics investigates numbers, laws, properties, possibilities, etc., as entities in their own right, since they seem to be presupposed by our very understanding of the scientific enterprise. The theory of abstract objects attempts to organize these objects within a systematic and axiomatic framework.[15]

So doing metaphysics results in lists of what there is which include abstract objects, as they are presupposed in scientific practice, plus an organization, an axiomatic framework. In the case of Zalta's axiomatic metaphysics, there is yet another particular purpose besides the purpose of a classification of objects presupposed by our understanding of the scientific enterprise: this is a way of doing metaphysics aimed at *not upsetting the naturalist* by admitting entities outside the spatiotemporal causal order, something which seems in many cases to be required by, or involved in, scientific explanation. This means among other things that Zalta – although pursuing a Meinongian project of metaphysics as a theory of objects – rejects or revises, in his own theory of objects, a number of Alexius Meinong's own proposals.[16] This is the case namely of (i) the idea that there are ways of *Sein* (*Existenz, Bestand*) and (2) the acceptance of both existence and non-existence of objects. Historically Meinong's options made room for his infamous florid ontology, which comprised real objects, abstract objects as well as non-existent objects. The ultimate reason for such a densely populated universe – the very antithesis of, say, the arid landscapes favoured by Quine – was the idea that all objects, whether they exist or not, possess the properties which are used to characterize them. Regardless of how naïve Meinong's *Gegenstandtheorie* might have been (this is Zalta's term at the beginning of his 1983 *Abstract Objects*[17] by which, to be fair, he mostly wants to contrast it with his own formal proposal), the fact is that with his theory of objects Alexius Meinong intended to fight what he saw as a prejudice. This was the prejudice of equating reality with *Wirklichkeit*,

[15] See Zalta 2004. See also Zalta 1983.

[16] As Dale Jacquette put it recently «Zalta's logic is more a logic of Platonic and Fregean abstract entities than of Meinongian beingless intended objects, that are neither actual nor abstract but ontically homeless.» (Jacquette 2015: 253).

[17] Zalta 1983 (Preface: xi).

i.e. actuality, in the sense of spatiotemporal and causally effective being. Meinong explicitly wanted to resist such prejudice. Many more things are real than what is spatiotemporal and causally effective, he thought. By itself this contrasts with Zalta's Neo-Meinongianism, since the fact is that, while sharing the classificatory outlook of Meinong's *Gegenstandtheorie* and the idea of metaphysics as involved in producing 'lists of what there is', Zalta equates existence with spatiotemporal existence. For Zalta, real, existing objects are concrete, actual, spatiotemporal objects, whereas non-existing objects are taken to be abstract.[18] Of course, what is at stake is not what we mean by 'exist' or 'exists': for Zalta 'existence' is a primitive theoretical notion that does not necessarily translate the English 'exists'. This is all very fine and explicit – I just want to call attention to the fact that it is against the background of such a conception of how to do metaphysics (understood as a pursuit intended not to upset the naturalist, which starts by equating being real with being spatiotemporal and causally effective, something which not only was not shared but was even opposed by Meinong himself), that fictional entities are taken to be *abstract* objects.

Naturally this is not a view of metaphysics one is forced to share. My intention here is not to argue against it but simply to point out that no concrete-abstract distinction is available independently of one's commitment on how to do metaphysics.[19] As Gideon Rosen puts it, the concrete-abstract distinction is a very important distinction, but there certainly is no standard account of how it should be drawn.[20] What we need then is a proposal about how the distinction might be fruitfully used.

---

[18]Zalta's commitments are very clear – we may take it that he thinks that in the twenty-first century we are all naturalists; at least this lies behind his conception of metaphysics. Yet need we be? Meinong himself did not at all share such inclination. One might say that opposing naturalistic and reductionistic inclinations philosophically lies at the heart of his *Gegenstandtheorie*.

[19]My central example of fictional entity in this article is Elizabeth Costello and I went through Cora Diamond's view of her nature as a character. Now Diamond's conception of how to do metaphysics is certainly very different from Zalta's. Diamond's term for her own conception of how to do metaphysics is 'realistic spirit'. 'Realistic' is to be contrasted here with philosophical 'realism'. It is closer to the use of the term outside philosophy. In Diamond's words «We may tell someone to "be realistic" when he is maintaining something in the teeth of the facts, or refusing even to look at them. (...) We also speak of realism in connection with novels or stories: and here again we have in mind certain kinds of attention to reality: to detail and particularity. (...) A further characteristic of realistic fiction, which is relevant in the same sort of way, is that certain things do not happen in it. People do not go backwards in time, pots do not talk, elves do not do chores while shoemakers sleep, and holy men do not walk unaided over the surfaces of lakes or oceans. We all know that if God sells wine in an English village, we do not call the story realistic (...). There is a third characteristic of realism outside philosophy, related to both of the others, and this is the significance of consequences, of causation. A man wanting to bring about some social reform, will be said to be unrealistic if he does not attend to how politics works.» (Diamond 1991: 40). Philosophical realism is taken not to be 'realistic' in Diamond's sense. Another important point of Diamond's realistic spirit is the idea that the nature and contents of a list of what there is is not what matters most for metaphysics discussions.

[20]Rosen 2014.

Let us go back to the fact that no concrete-abstract distinction is available independently of one's commitment on how to do metaphysics. From the viewpoint of the history of philosophy there is yet another interesting fact. There is no way to avoid noticing how recent the concrete-abstract distinction at play e.g. in Zalta is. Let us say that we, quite naturally, think that the concrete-abstract distinction is close to Plato's distinction between Forms and the Sensible. After all, commitment to abstract entities is very often identified as Platonism. A problem arises immediately: Plato's Forms or Ideas are not causally inert. Forms are in fact supposed to be the very causes of existing particulars such as this chair, or this horse. Platonic Forms may be non-spatiotemporal but they are certainly not abstract in anything like the modern sense of 'abstract' as causally non-efficacious.

As an alternative, we may try to trace the concrete-abstract distinction to classic empiricism. There we find a distinction between concrete and abstract ideas that seems be closer to grammar: we speak e.g. of concrete and abstract nouns – say, *white* and *whiteness* (*o branco* and *a brancura*, in Portuguese). Rosen notices this is the case with e.g. John Locke in his *Essay Concerning Human Understanding.*[21] Speaking of general ideas –abstract and harder for the mind to come by, according to him – Locke mentions the general idea of a triangle, a triangle that is neither Rectangular nor Scalenon. Such idea of a triangle is abstract in Locke's sense. We would expect that something like this concrete-abstract distinction for ideas, in Locke's sense, would be held by all empiricists. But was it? It was certainly not held by Berkeley or Hume: they actually disagreed with Locke on this. For Berkeley, there is not such thing as an abstract idea, i.e. there is no general idea of a triangle that is neither Rectangular nor Scalenon. Any idea in a mind is for Berkeley a particular with particular properties: Berkeley's concretism about ideas in fact underlies his (metaphysically immaterialist) empiricism. And when a twentieth century empiricist such as Quine comes to regard abstract and universal as equivalent this is certainly very different from what early empiricists took abstractness to be. In other words, if we look at the history of philosophy looking for origins of the concrete-abstract distinction, Platonism might seem a source of our conception of abstractness, but at least historical Plato is not. If we look at the history of empiricism, as another possible source of our conception of abstract we see that being an empiricist does not *per se* settle any particular view of the concrete-abstract distinction. Perhaps in the twenty-first century we are all supposed to be naturalists, yet this does not *per se* give us a concrete-abstract distinction to work with. The question is simply not settled.

The question to ask then is: if there is no straight association of the concrete-abstract distinction with historic Platonism, or classic empiricism, where does our standard view come from? Where does the idea that an object is abstract if and only if it is non-spatiotemporal and causally inefficacious come from? This question does have an answer. One suggestion common to Jocelyn Benoist (Benoist 2005) and Gideon Rosen (Rosen 2014) is that historically it comes from the discussions

[21] Locke 1959, IV, Vii, 9.

at the origins of phenomenology and analytic philosophy, discussions regarding the objectivity of thought and judgment (e.g. arithmetic judgment). It belongs thus in a quite recent, nineteenth–twentieth century, context. Thoughts and judgments are neither material particulars like this chair or that computer nor (for early phenomenologists and analytic philosophers) psychological processes. We would like to take it that they are objective – but what objectivity is this? Accounting for what they saw as the objectivity of judgments, thoughts or propositions was a problem common to people like Bernard Bolzano, Franz Brentano, Edmund Husserl, Kazimierz Twardowski or Gottlob Frege. This is the galaxy of authors in which Alexius Meinong's work belongs (to which one should add Bertrand Russell, who famously derided Meinong in his 1905 *On Denoting*). This is the context where our current talk of something as abstract if it is both non-physical and non-mental started. It is important to notice though that what is being discussed here is objectivity (of thoughts, or propositions), not objects in a list of what there is. The reason why it is historically illuminating to think of Alexius Meinong as belonging in such a constellation of authors is that he is very peculiar about one thing. The others authors mentioned do not see their research into the objectivity of thought as necessarily resulting in a *Gegenstandtheorie*, a theory of objects. This is not the sole way to pursue an investigation into the nature of objectivity– it is simply Meinong's way. In alternative one could take e.g. Frege's way, and speak of objects only much later, as it were, i.e. when much more is in place regarding a view of language and thought. We may even argue that Frege had a more 'abstract' view of 'object' – a view not connected with our sensorial experience or the paradigmatic character of a spatiotemporal particular within a conception of reality as *Wirklichkeit* but rather a view connected with the structure of that which we are capable of thinking and saying, conceived in terms of functions and objects.[22]

12.4 Conclusion: Sketch of Argumentation for a Philosopher with a Different Temperament

So back to square one when thinking about fictional entities. We are bound to think of them as abstract in a particular sense if we have independent motivation for pursuing metaphysics a certain way ('metaphysics as lists of what there is') and for being a certain kind of naturalist. But let us say that we do not have such independent motivation. I want to finish by sketching an argument for fictional entities being in a specific sense concrete which might appeal to someone with very different philosophical inclinations from those of Meinong or Zalta. Someone who sees the pursuit of metaphysics with totally different eyes. Since I took my central example in this essay from Cora Diamond, let us say it is her.

[22]Of course this is not the usual view of Frege. I am taking it from Charles Travis (Travis 2013).

The argument starts from a parallel between characters and people as particular spatiotemporal entities and goes the following way. Take a spatiotemporal particular such as myself (M). Whether M is F depends on whether it can be brought under a generality (being F). That depends on everything the particular case is. So I give a particular description of myself (e.g. that I am Portuguese, that I am now sitting, that I am now writing, that I am wearing a black sweater, that I am hungry, that I will shortly be teaching a course on moral philosophy, that just before working on this paper I was reading Kafka's *Brief an dem Vater*). Make the description more Jorge Luis Borges-like, a very long and very detailed sequence of descriptions of me. What is the status of such descriptions of a particular spatiotemporal object such as myself? Me being me here and now falls under various generalities; yet what I am representing as being a certain way when I think of myself as falling under such generalities is not reducible to a certain list of properties that I have. How would it be? Typical concrete objects like me and you have a lot of accidental properties. In this they contrast with preconceived objects have no extra properties beyond their essential.

What about entities in fiction? Need we think of entities in fiction such as Elizabeth Costello as fully preconceived objects, exhausted by descriptions at time t? Not necessarily. Yet this is certainly taken for granted in many discussions. In fact, in most discussions of fictional entities in the literature fictional entities are like characters as described in a bad novel or film, i.e. they are reducible to a certain list of properties. Somebody, i.e. such character, is thus and so, we are told – and he/she is just as it is described in a novel. No more properties. But precisely, going back to my initial example of the Singer-Diamond dispute about moral thinking, in a good novel, a character described as being a certain way does not reduce to the descriptions of it presented within the novel. That is why it is worth it to read the novel, or play, over and over, in contrast with being told 'Hamlet is the Prince of Denmark, son of the recently deceased king, nephew of the new king, his father's brother and successor, who married his mother, and he wants to kill his uncle but has qualms about it'. If you think this is sufficient, then you may go home; you do not have to watch the play performed (or read it). Or let us say that you say (about Coetzee's novel) 'Elizabeth Costello is a writer, vegetarian and a defender of animal rights' and then you replace Elizabeth Costello with an argument in favour of animal's rights. The job is done – you do not have to read the book. If we think this way the aesthetic, moral and cognitive practice involving fictional entities is left totally unaccounted for. Why do we even do that (e.g. watch plays or read novels)? The value of doing it, cognitive value included, is left unaccounted for. We could simply wait and be given descriptions such as the ones above. But what has to be accounted for is our engaging with such works resulting in seeing things anew. Seeing things anew is what is taking place in our moral thinking when we engage with works such as Hamlet, or *The Lives of Animals*, or Elizabeth Costello.[23] In the case of Elizabeth Costello her being unplain is that with which our moral thinking

[23]And then I would say: thinking about ourselves, in this world, the only world, following the directives of fiction. This is where the philosophy of language questions regarding fiction may be posed again.

engages with. This happens in a way not accounted for by argument – argument works only once the work of conceiving has been done. And if that work were done for good we would not need literature, or fictional entities, at all.

My suggestion then is that the notion of concreteness may help us figuring out what is taking place here. The notion of concreteness does indeed apply first to things in the environment – this tree, this table, myself. But why should we take it to be thus exhausted? Questions regarding concreteness, as questions regarding abstractness, have to do with not only with how a thing is but with how we think of it – any thing – this tree, Hamlet, Elizabeth Costello or numbers and I are. It is, thus, more general.

So, one thing that saying fictional characters are concrete may mean is that one does treat them like we treat (concrete) things in an environment. This is not the point that C-Meinongians stress; they see entities in fiction as non-existing concrete entities because they are correlates of exactly the same properties which are predicated of common objects. My point is rather that we think of them, and predicate them, not thereby needing to think they reduce to what they are said to be in such thinkings and predications (as is the case of our dealings with concrete things in our environment). From this I believe we get an explanation regarding why and how fictional characters matter for our moral thinking (and thus for moral philosophy) which makes the shape of the problem of fictional entities different in the eyes of the philosopher. Taking a fictional character in a novel to be a subsisting abstract entity exhausted by a list of properties has no further explanatory purpose. It ends there; it does not do any work e.g. in moral philosophy and in making sense of our moral practices. It leaves us empty-handed. But what is at stake in the discussion of fiction and fictional entities is not just the subsistence of objects or their original creation (which is what artefactualists, possibilists and Neo-Meinongians discuss[24]). What is at stake is the nature of our thinking while we engage with the works as well as the openness of such engagement, i.e. the fact that it keeps producing novelty. This is what needs to be considered and a way of considering it is what we gain by taking fictional entities to be concrete. We want to understand how our confrontation with the character Elizabeth Costello results in, how it changes our ways of thinking in ways that could not be fully predicted or intended by Coetzee, the author. This is where thinking of Elizabeth Costello and her unplainnesss as concrete, helps. The question is not just what there is, whether it was created or not, as it figures in a list of things there are.

Since I gave myself as one example of a typical concrete object, now consider the following: one might say that I am completely here, now.[25] This is of course a very interesting claim, as well as a very hugely loaded one, metaphysically speaking. It is supposed to contrast with the fact that fictional entities, like Elizabeth Costello, are incompletely described, incompletely determined. But in which sense am I

---

[24] A further difficulty here is that the creation might be diffuse, involving many authors and acts of creations, as Thomasson admits (Thomasson 1999: 165, n. 3).

[25] I thank Concha Martínez for this observation.

completely here (now)? In the sense of full determination of my predicates? Am I really determinate down to the last detail in this actual world in which I exist? In order for that to be the case, there would have to be something like a Leibnizian complete notion of an individual, me, SM – the unfolding of all my past, present and future predicates would have to be available from the view point of God (at least this is how Leibniz once conceived of the complete notion of an individual). Do we want to say that? If not, why not say that both I and Elizabeth Costello are concrete in the sense of being *further thinkable?* In that sense Elizabeth Costello, or Hamlet, are not that different from me. This is one suggestion regarding the nature of fictional entities.

The additional advantage is that thinking of fictional entities as concrete in this sense of being further thinkable (in *that* there is no difference between Hamlet and me) does lots of work in philosophy; in particular it leads us somewhere in moral philosophy. We know e.g. what is wrong with a position such as Singer's about the nature of moral thinking, according to when it comes to its moral 'purpose', fiction might simply be replaced by rational argument, and fictional characters might simply be taken to be lists of properties. According to him this is sufficient to account for our moral thinking when we engage with fiction. But it is not. One reason it cannot be thus replaced is because fictional characters are in the sense I am proposing, concrete.

Such way of thinking of entities in fiction is meant to call attention to an aspect of moral thinking which is orthogonal to that which currently divides moral realists and moral fictionalists: what is at stake here is not so much the nature of our moral commitments and whether they require true representation of a particular domain of fact but rather the novel conceivings of moral matters, as they take place in our engagement with fiction. That is what we then, say, in arguments, may then eventually take to be (moral) facts.

References

Benoist, J. (2005). Esquisse d'une théorie non-intentionnaliste de la fiction. In *Les limites de l'intentionalité – recherches phénoménologiques et analytiques*. Paris: Vrin.

Coetzee, J. M. (1999). *The lives of animals*. Princeton: Princeton University Press.

Coetzee, J. M. (2003). *Elizabeth Costello*. London: Vintage, Random House.

Diamond, C. (1991). *The realistic spirit: Wittgenstein, philosophy, and the mind*. Cambridge, MA: The MIT Press.

Diamond, C. (2008). The difficulty of reality. In S. Cavell, C. Diamond, J. McDowell, I. Hacking, & C. Wolf (Eds.), *Philosophy and animal life*. New York: Columbia University Press.

García-Carpintero, M. (2016). *Relatar lo ocurrido como invención – una introduction a la filosofía de la ficción contemporánea*. Madrid: Cátedra. Teorema.

Jaquette, D. (1996). *Meinongian logic. The semantics of existence and nonexistence*. Berlin/New York: De Gruyter.

Jaquette, D. (2015). *Alexius Meinong: The shepherd of non-being*. Berlin: Springer.

Lewis, D. (1986). *On the plurality of worlds*. Oxford: Blackwell.

Locke, J. (1959). *An essay concerning human understanding*. New York: Dover.

Miguens, S. (2011). 'What some philosophers wouldn't dream of considering as part of their job': Cora Diamond on ethics and human lives. *Revista Portuguesa de Filosofia, 67*(I), 155–167.
Miguens, S. (2017). Onde nos leva Wittgenstein em filosofia? As respostas de Elizabeth Anscombe e Cora Diamond. In B. Fernandez, L. Veríssimo, & D. Couto (Eds.), *Por Que Razão? Razões para acreditar, para agir e para preferir*. Porto: FL-UP, MLAG Discussions Papers 8.
Miguens, S. (2019). *Uma leitura da filosofia contemporânea – Figuras e movimentos*. Lisboa, Edições 70.
Parsons, T. (1980). *Nonexistent objects*. New Haven: Yale University Press.
Rosen, G. (2014). Abstract objects. In E. Zalta (Ed.) *The Stanford encyclopedia of philosophy* (Fall 2014 Edition). http://plato.stanford.edu/archives/fall2014/entries/abstract-objects/
Salis, F. (2013). Entidades ficcionais. In J. Branquinho & R. Santos (coords.) (Eds.), *Compêndio em Linha de Problemas de Filosofia Analítica*. Lisboa: Centro de Filosofia da Universidade de Lisboa.
Thomasson, A. (1999). *Fiction and metaphysics*. Cambridge: Cambridge University Press.
Travis, C. (2013). *Perception – Essays after Frege*. Oxford: Oxford University Press.
Voltolini, A. (2015). A suitable metaphysics for fictional objects. In S. Brock & A. Everett (Eds.), *Fictional objects* (pp. 129–146). Oxford: Oxford University Press.
Zalta, E. (1983). *Abstract objects – An introduction to axiomatic metaphysics*. Dordrecht: Reidel.
Zalta, E. (2004). The theory of abstract objects. In *Homepage of the metaphysics research lab*. https://mally.stanford.edu/

Chapter 13
Abstract Objects and the Core-Periphery Distinction in the Ontological and the Conceptual Domain of Natural Language

Friederike Moltmann

Abstract This paper elaborates core-periphery distinctions in the ontological and the conceptual domain of natural language. The core-periphery distinction is essential for the pursuit of natural language ontology and has in fact been made implicitly by any philosopher present or past when appealing to natural language for motivating an ontological notion or view. The distinction plays a central role in the main thesis of my 2013 book *Abstract Objects and the Semantics of Natural Language*, that natural language permits reference to abstract objects only in its periphery, not its core. The paper explores how the core-periphery distinction relevant for ontology appears to be structurally anchored and relates to the more familiar core-periphery distinction that Chomsky drew for syntax.

Keywords Abstract objects · Core-periphery distinction · Propositions · Properties · Numbers · Degrees · Functional categories

13.1 Introduction

Abstract objects such as properties, numbers, and propositions have been a topic of philosophical controversy since antiquity. While there are a range of philosophical arguments for and against abstract objects, philosophers have also appealed to natural language, generally arguing that abstract objects are well-reflected in natural language, in what appear to be abstract terms, such as *happiness*, *the number of planets* or *that*-clauses. This has given rise to the widespread view that natural language involves a rich and philosophically controversial ontology of abstract objects.

F. Moltmann (✉)
CNRS, Paris, France
e-mail: fmoltmann@univ-paris1.fr

J. L. Falguera, C. Martínez-Vidal (eds.), *Abstract Objects*, Synthese Library 422,
https://doi.org/10.1007/978-3-030-38242-1_13

A systematic linguistic investigation, based on a much greater range of data, whether and how natural language (or rather English) involves an ontology of abstract objects was the aim of my 2013 book *Abstract Objects and the Semantics of Natural Language*. The book rejects the common view that natural language permits pervasive reference to abstract objects and instead endorses what I will now call *the Abstract Objects Hypothesis*:

(1) The Abstract Objects Hypothesis

Natural language does not involve reference to abstract objects in its ontological core, but only in its ontological periphery.

The book, for example, argues that *happiness* is not a term standing for an abstract property object, but rather a plural term, referring to all the particular happiness features (tropes) at once. *The number of planets*, the book argues, does not stand for an abstract number, but rather for a number trope (the numerical aspect plurality of the planets).[1] *That*-clauses moreover are not considered referential terms at all standing for objects.

Crucial in the formulation of the Abstract-Objects Hypothesis is the relativization to a core-periphery distinction. This is an important distinction that has always been made implicitly when philosophers appealed to natural language in support of a philosophical view, but has hardly ever been articulated and theoretically elaborated. Whenever philosophers (and linguists) appeal to natural language to motivate an ontological notion or view, they are careful to only draw on expressions from the ontologically relevant core of language ('the ontological core', for short), not its periphery. Thus, it is considered legitimate to appeal to the existence in English of expressions like *happiness, the number of planets,* and *that it is raining* when motivating properties, numbers, and propositions as abstract objects, but not expressions like *the property of being happy, the number eight*, and *the proposition that it is raining*, that is, abstract terms particular to philosophical discourse, but also legitimate for use by any competent speaker of English.

Intuitively, the ontological core consists in expressions or uses of expressions not involving ontological reflection, whereas the ontological periphery consists of expressions or uses of them involving ontological reflection. This distinction between core and periphery reconciles the fact that only certain sorts of expressions or uses of expressions are considered indicative of the ontology implicit in natural language with the fact that natural language contains expressions that clearly can be used to make reference to abstract objects of various sorts and permits the introduction of new expressions (or uses of expressions) for that purpose.

An important feature of the core-periphery distinction relevant for natural language ontology is that there appear to be significant constraints as to what parts of language can lead to a use in the periphery. The ontological core-periphery distinction appears to be structurally anchored, roughly going along with the lexical-functional distinction, with full nouns freely being able to lead to a use in the

[1] For the notion of a trope as a particularized property see Williams (1953).

periphery but not morphological features or functional categories. This means that the ontological core-periphery distinction is not just a somewhat elusive distinction based on mental acts or attitudes, but is also anchored in the structure of language.

Natural language ontology with such a core-periphery distinction goes along with a particular conception of ontology. First, the ontology reflected in natural language should be viewed not as an ontology of what there really is but rather an ontology of appearances (Fine 2017), in the sense of consisting of objects that serve as values of referential noun phrases and bear a range of properties, but that may be derivative or merely conceived objects. Second, the ontology reflected in language should not just consist of a single domain of entities and their associated categories, but should allow for considerable flexibility and expansion of the domain, by including ontological operations introducing new entities.

This sort of expandability matches that of the lexical-conceptual domain of language, which displays a similar core-periphery distinction. Thus, it has been argued that only certain concepts allow for 'conceptual engineering' (Eklund 2015, Chalmers 2011), and again, there appear to be structural constraints to what extent meanings can be modified or expanded on non-ordinary, philosophical uses of language.

Core-periphery distinctions have also been made for syntax (Chomsky 1981, 1986, 1998) and for phonology (Itô and Mester 1995a, b). While the core-periphery distinction that Chomsky drew for syntax may in some ways coincide with the one relevant for ontology, the core-periphery distinction in phonology is an entirely different one.

The overall aim of the paper is to elaborate the ontological core-periphery distinction in the context of the cognitive ontology reflected in natural language, focusing on apparent reference to abstract objects. The paper will start by presenting standard arguments for an against abstract objects. It will then lay out the background of natural language ontology and in particular the notion of a language-related, constructional ontology it involves. The main part of the paper consists in elaborating core-periphery distinctions for the ontological and the conceptual domains of natural language, followed by a brief discussion of core-periphery distinctions that have been made in generative syntax and phonology. Finally, the paper gives a concise presentation of the reanalyses in Moltmann (2013a) of apparent abstract terms in the core of language and proposes a way to make sense of the semantics and ontology of abstract terms in the ontological periphery. In an appendix, the paper discusses apparent problems for the Abstract Objects Hypothesis.

13.2 Abstract Objects, Natural Language, and the Core-Periphery Distinction

There are a range of philosophical motivations for abstract objects of the various sorts as well as philosophical arguments against them. Properties have been considered the basis for relations of resemblance and the individuation of objects; propositions have been regarded important as the sharable contents of thought; and numbers, sets and other mathematical objects are obviously central in the ontology of mathematics. Yet abstract objects are also considered problematic in that they are not part of the empirical world and in particular cannot enter causal relations and act as the objects of perception, raising the issue of how we can have epistemic access to them. Moreover, it is unclear how abstract propositions can be grasped and act as the content of mental attitudes. In addition to providing metaphysical and epistemological arguments for and against abstract objects, philosophers have also drawn on natural language, generally to give support for abstract objects (but also sometimes to argue against them).

In analytic philosophy, this is particularly the case in the work of Frege and Neofregeans (Wright 1986; Hale 1983), who take the syntactic category of referential noun phrases (singular or referential terms) to be indicative of objecthood ('an object is what a referential term may stand for'). Properties, propositions, and numbers then are generally considered objects as there appear to be referential noun phrases for them (*the number of planets / eight, wisdom, that*-clauses like *that it is raining*). This would extend to pure quotations (as in *'Red' means 'rouge'*), which are commonly taken to involve the formation of referential terms standing for expression types. The putative involvement of objects is not limited to the role of semantic values of referential terms. Objects may also just act as implicit arguments of predicates. For example, degrees are abstract objects that are commonly regarded as implicit arguments of gradable adjectives.[2]

Given apparent abstract terms such as *wisdom*, *the number of planets*, and *that*-clauses, it has become a widespread view that natural language involves pervasive reference to abstract objects, a view which led some philosophers skeptical about abstract objects to reject language as a guide to ontology.

In my own work (Moltmann 2013a), I have rejected this common view about the involvement of abstract objects in the semantics of natural language. Instead I have argued for the view that natural language does not involve reference to abstract objects in the relevant part of natural language, its ontological core, but only its periphery (the Abstract Objects Hypothesis). Relevant referential expressions that were generally considered abstract terms are now analysed differently. Instead of referring to abstract objects, they are either no longer considered referential terms or else are considered terms involving reference to something other than single abstract objects, say, pluralities of particulars, particularized properties (tropes), or variable objects with concrete entities as manifestations.

Of course, there are natural language terms that cannot be regarded other than as terms standing for abstract objects, such as *the property of being wise, the number eight*, *the proposition that it is raining*, and various abstract terms that philosophers may be using. But such terms will be part of the ontological periphery, not the ontological core of language (see Sect. 13.4).

The core-periphery distinction is indispensable when investigating the ontology reflected in language, and in fact it is a distinction that has been presupposed by

[2]Degrees can also be made explicit as in *ten meters tall*.

all philosophers throughout the history of philosophy that have appealed to natural language in support of a philosophical view. Thus, Frege and Neo-Fregeans knew to stay away from terms like *the property of being wise, the proposition that it is raining*, and *the number eight,* being aware that they are unsuited to make their point. Instead they drew on expressions like *the number of planets* and *eight* (Frege 1884), *wisdom* (Hale 1983), and *that*-clauses (all philosophers arguing for propositions since Frege 1918/19). Philosophers of course also stay away from expressions specific to philosophical discourse ('technical terms') and philosophical (that is, non-ordinary) uses of natural language expressions whose semantic values come from a particular philosopher's domain. Without a core-periphery distinction the ontology described by any philosophers' theory would be part of the ontology of natural language, undermining the pursuit of natural language ontology entirely.

It is an undisputable fact, though, that natural language permits the introduction of expressions or uses of expressions meant to have semantic values within a particular philosopher's ontological theory (despite attempts by ordinary language philosophers to sanction such uses). Thus a philosopher can introduce a particular notion of existence, truth, or parthood and use the nouns *existence*, *truth*, and *parthood* to be associated with such a notion rather than whatever the ordinary meaning of the expression may be from which such nouns are derived (*exist, true, part*). Such uses are not illegitimate, but rather are made possible by a fundamental feature of language or rather the conceptual and ontological structure going along with it, namely what I will call *expandability*. Expandability consists in the possibility of conceptual modification or ontological expansion that goes along with a non-ordinary use of natural language expressions, with the introduction of new expressions, or with particular constructions already present in the language (such as constructions of the sort *the property of being happy*).

Expandability of the conceptual and ontological domain associated with language allows for a form of creativity in more of the standard understanding of the term than the creativity of language use in the Chomskyan sense, which consists in the ability of a speaker of a language to produce and understand an indefinite number of new expressions against the background of limited experience. Creativity in the conceptual and ontological domain manifests itself in the introduction of new concepts or objects based on reflection (as opposed to implicit acceptance).[3]

13.3 Natural Language Ontology

The following is a brief outline of the discipline whose subject matter is the ontology of natural language, natural language ontology.

[3] 'New' in the case of objects in the ontological domain means objects that are not yet part of the domain of interpretation of the language, whether the objects are real or merely conceived. See Sect. 13.2.

With its referential noun phrases, which take entities as semantic values as well as with its lexical predicates and constructions that involve entities in other ways, natural language reflects an ontology. This ontology is the subject matter of a particular branch of metaphysics, that of natural language ontology. More specifically, natural language ontology is part of descriptive metaphysics in Strawson's (1959) sense or what Fine (2017) calls 'naïve metaphysics'. Descriptive metaphysics has as its subject matter the ontology reflected in our ordinary judgments. Natural language ontology has as its subject matter the ontology reflected in linguistic intuitions, that is, judgments about the acceptability or grammaticality of natural language sentences and constructions.

What is important about descriptive metaphysics is that it is not about the ontology of what there really is. This is instead the subject matter of a different branch of metaphysics, what Fine calls 'foundational metaphysics'.[4] Descriptive metaphysics and natural language ontology in particular concerns itself with how things appear, given the data, without addressing the question of whether they are real (which is to be addressed only by foundational metaphysics). For natural language ontology; this means that no foundationalist consideration should come into play when positing objects as semantic values, such as assumptions as to whether those objects really exist (in the sense of being fundamental) or what they may be reduced to. More important is what sorts properties the semantic values of referential noun phrases may have, as is reflected (at least to an extent) in the applicability of types of natural language predicates. The domain of objects in the ontology of natural language thus may include merely conceived objects besides objects that happen to be actual ones. (This will also be important for how to make sense of the expandability of the ontological domain of the language through technical or philosophical discourse in Sect. 13.4.)

The subject matter of natural language ontology is not the ontology that ordinary speakers (non-philosophers) naively accept when thinking about what there is.[5] The latter is the subject matter of folk metaphysics, not natural language ontology. What speakers accept when they reflect does not matter for natural language ontology. Natural language ontology rather deals with the ontological categories, notions, and structures that are implicit in language whether or not speakers would accept them upon reflection. The ontology of natural language thus is better to be characterized as the ontology that a speaker *implicitly accepts* when using the language and as such is distinguished from both the reflective ontology of ordinary speakers as well as philosophers and the ontology of what there really is (Moltmann 2017, 2019).

Natural language ontology is not just a new discipline (as a branch of both metaphyics and linguistics), but in a way was pursued by philosophers throughout

[4]For Strawson (1959), descriptive metaphysics rather contrasts with what he calls 'revisionary metaphysics'. The aim of revisionary metaphysics, for Strawson, is to conceive of a better ontology than the one we ordinarily accept. (Strawson does not specify further how 'better' is supposed to be understood.)

[5]Fine's (2017) term 'naïve metaphysics' thus is misleading, and 'descriptive metaphysics' a better term to use for the branch of metaphysics that comprises natural language ontology.

the history of philosophy whenever philosophers drew on natural language in support of a metaphysical notion or view. Philosophers when appealing to natural language, though, appealed to only certain types of linguistic data and not others, and thus implicitly followed a particular methodology, the same that also guides natural language ontology as an emerging contemporary discipline that is part of both descriptive metaphysics and linguistics (semantics). For example, metaphysical statements of the sort *there are properties, there are numbers* etc. are not considered indicative of the ontology reflected in natural language. By contrast, sortal presuppositions of lexical predicates, for example, are.

An important feature of the ontology of natural language is that it involves not just a particular domain of objects and their associated ontological categories, but ontological operations that serve the introduction of new objects. In that sense it is a *constructional ontology* (Fine 1991). To an extent, the ontological operations go along with the compositional semantics of particular syntactic constructions. Here are some examples (though there are no unanimous views as to the semantics they involve). First, the introduction of a trope (particularized property) goes along with the construction NP's N_A, where N_A is the nominalization of the adjective A, as in *John's happiness, Socrates' wisdom, Mary's courage* (Moltmann 2013a). On a common view, formation of a kind goes along with bare (determinerless) plural and mass nouns in English (*giraffes are rare, water is transparent* Carlson 1977). On an equally common view, the construction of definite plurals (*the students*) goes along with the introduction of a sum composed of the individuals that fall under the corresponding singular noun (Link 1983).[6] Finally, definite NPs with a functional noun as head or modified by an intensional relative clauses go along with the introduction of a variable object, an object that has potentially different concrete manifestations at different times and perhaps (possible) situations. Examples are *the president of the US* (as with *is elected every four years*), *the water in the pool* (as with *decreased*), and *the book John needs to write* (Moltmann 2013a, to appear). The linguistic ability to use those constructions goes along with the implicit acceptance of the relevant ontological operation, and an actual use goes along with the implicit acceptance of the output of the operation.

In addition to syntactic constructions, there are syntactic categories and features that convey ontological categories (to give two simplified examples: verbs with events, adjectives with tropes), and syntactic knowledge of them goes along with the implicit acceptance of those ontological categories.

The constructional ontology reflected in natural language involves more differentiated forms of ontological acceptance. There will not just be a single notion of acceptance, acceptance of the entities in the domain of the ontology. There will also be a notion of acceptance of ontological operations, and that with or without acceptance of the output of such operations. Implicit acceptance of an ontological operation then needs to be distinguished from the reflective acceptance that may or

[6]In the latter two cases, there are also arguments, however, that the NPs plurally refer, rather than leading to the composition of a new entity (Moltmann 2013a; Yi 2005, 2006). See also Sect. 13.6.

may not apply to the output. Such reflective acceptance would then lead from the ontological core to the periphery of language, as will see.

13.4 The Core-Periphery Distinction in Ontology and in Semantics

13.4.1 The Core-Periphery Distinction in Natural Language Ontology

An important distinction differentiating between linguistic data indicative of the ontology of natural language and data not indicative of it is the distinction between the ontological core and periphery of language. The distinction is reflected in the sorts of data that contemporary philosophers or linguists know to choose when they pursue natural language ontology, and it is reflected in the sorts of data that philosophers throughout history knew to choose when they in fact engaged in natural language ontology. It is a distinction that is presupposed just as much by the appeals to natural language that philosophers considered appropriate throughout history and by contemporary pursuits of natural language ontology by linguists and philosophers. It is thus a distinction that is central to natural language ontology as a historical practice and an emerging discipline.

Most importantly, full nouns conveying ontological categories (sortals) are generally not taken to be indicative of the ontology of natural language. The fact that nouns like *property, proposition, number, trope, kind, sum*, etc. can be used, legitimately, in English to refer to entities and quantify over them is generally not taken as evidence that the ontology implicit in language includes objects of such categories. Moreover, what I call 'reifying terms', noun phrases that are formed with sortal nouns like *the number eight, the proposition that it is raining*, or *the truth value true*, have generally not been taken to be indicative of the ontology of natural language.[7] Thus, Frege (1918/19) did not motivate numbers as objects appealing to the presence of the construction *the number eight*, and Frege (1918/19) did not motivate truth values as objects by appealing to *the truth value true* (but rather by considerations regarding embedded sentences and the sense-reference distinction across categories). Yet, reifying terms with a selected set of sortals are fully productive and thus lead to referential terms standing for objects, the sorts of objects that should fall under the sortal that is part of the construction. What then is the difference to the constructions discussed in the last section, which involve ontological operations yielding composite or derivative objects (tropes, kinds, pluralities, variable objects) that would fall under suitable sortals? In both

[7] See Moltmann (2013a, Chapter 6) for a discussion of 'reifying terms'. Reifying terms divide into close appositions such as *the number eight* and *the truth value true* and constructions of the sort *the proposition that* S, *the fact that* S, *the property of being wise.*

cases we have complex syntactic constructions, but only in the second case will the objects that are the output of the operation count as part of the ontology of natural language. The reason is the involvement of a sortal in the first case, which makes its output not be part of the ontology implicit in language. In the second case, the outputs are *implicitly accepted* by anyone making use of the construction. In the first case, the construction itself involves explicit recognition of the output as an object falling under the respective sortal, and thus an act of reflection. Even if the construction itself with the ontological object-introducing operation is part of the core of language (and thus is implicitly accepted), the use of the construction with its involvement of an act of ontological recognition (or reflection) makes the output be part of the periphery, and thus not part of the implicitly accepted ontology.

The feature driving the ontological core-periphery distinction is thus ontological reflection: expressions (or uses of expressions) belong to the ontological periphery if their use involves some degree of ontological reflection, as would go along with the use of a sortal.

Note that the core-periphery distinction is not a strict, but a somewhat gradual distinction. While sortals like *set, sum, trope*, and *proposition* will require not just categorial recognition, but some technical or philosophical knowledge, this is presumably not so for *fact* (*the fact that* S) or *state* (*the state of John being happy*).[8]

The reflection that is characteristic of the periphery goes along with the use of full sortal nouns, which involves the explicit recognition of an object as belonging to a particular ontological category. By contrast, there are linguistic categories that convey ontological categories, but do not involve reflection and do not permit conceptual modification on a non-ordinary use. An example is plural morphology. Whatever notion of plurality it may in fact involve, that notion could not possibly be altered in a context of use to accommodate a particular philosophical view of plurality. Also the object-introducing constructions mentioned in the last section are of that sort, that is, definite NPs that introduce tropes, bare plurals and mass nouns that introduce kinds, and definite NPs that introduce variable objects. The use of those constructions does not permit altering the way the so introduced objects are conceived.

The ontological periphery comprises particular construction types, expressions, and uses of expressions (as well as what they stand for). As such it also includes ‘technical terms’, expressions introduced within a particular philosopher’s theory. It includes certain sorts of uses of expressions, namely just the non-ordinary, ‘philosophical’ uses of expressions that have been the subject of critique of ordinary language philosophy (Wittgenstein, Ryle, Malcolm). Rather than dismissing such uses as illegitimate, from the present perspective they do have their proper place, relying on a legitimate expandability of the semantics and ontology associated with

[8]There are also other sorts of terms that ordinary speakers use that appear to refer to abstract object. One example are number terms of the sort *the eight*, which clearly are terms referring to numbers as abstract objects, but certainly are used by ordinary speakers, even children (though such number terms are rather limited, just up to *the twelve* in my language). (*The eight* arguably is also a close apposition like *the number eight*, but involving a silent noun NUMBER.)

language. If expressions are used in a non-ordinary way to convey metaphysical notions, then those notions will be part of the ontological periphery of the language, not its core.

One issue the ontological core-periphery distinction raises is whether the core should represent the universal part of the cognitive ontology reflected in natural language as opposed to an ontology that would be specific to particular individuals, groups of individuals, or points of views. We have seen that part of the (core) ontology of natural language goes along with the semantics of particular syntactic constructions. There is no particular reason why syntactic constructions that serve the composition or introduction of a new entity should be shared among all languages. However, it should be only in the ontological core of languages that universally shared features of a cognitive ontology can be found.

13.4.2 The Core-Periphery Distinction in Semantics and Philosophical Conceptual Theory

While the core-periphery distinction is indispensable for pursuing natural language ontology, it is a distinction that may be drawn and has been drawn also for other domains associated with language, in particular lexical/conceptual meaning, syntax, and phonology. In what follows, I will briefly discuss the distinctions in those other language-related domains.

Like the ontology of natural language, lexical meanings in natural language do not form a rigid domain, but rather permit a great range of flexibility and polysemy, which has given rise to theories according to which the lexicon includes operations on meanings generating other meanings (the theory of the Generative Lexicon of Pustejovsky 1995). The flexibility that lexical meanings display also consists in the possibility for a language user to precisify or otherwise modify a given lexical meaning. When a philosopher engages in such modification of conceptual meaning for philosophical purposes, this is what is called ‘conceptual engineering’ in recent philosophical discussions.[9]

Engaging in conceptual engineering is not a privilege for philosophers, of course; the very same sort of non-ordinary use of expressions and introduction of a new or modified use of an expression is available to any competent speaker of the language.

Like ontological operations that lead from the ontological core to the periphery, conceptual engineering is a legitimate operation of expansion of the domain of conceptual meaning. It thereby also gives rise to a core-periphery distinction. Let me call this *the semantic core-periphery distinction*, a distinction roughly between what one may call ‘ordinary meaning’ and ‘non-ordinary meaning’.

[9]Conceptual engineering has been advocated as a replacement of the standard approach in analytic philosophy, conceptual analysis (Cappelen 2018).

There may be another, more important core-periphery distinction in the conceptual domain, based on potential limits to conceptual engineering (Chalmers 2011; Eklund 2015; Cappelen 2018). Concepts that resist conceptual engineering have been discussed as 'bedrock concepts' by Chalmers (2011) and as 'conceptual fixed points' by Eklund (2015). These are concepts so fundamental that they permit no modification in the context of a particular philosophical view. A distinction between a core consisting of such concepts and a periphery would define a fairly wide periphery of conceptual meaning. In such a distinction, which I will call *the conceptual core-periphery distinction*, the core would be a universal part of the human conceptual system, which would match a potential universal part of the core ontology of a given natural language.

Limits to conceptual engineering do not just concern concepts as such. Limitations to legitimate modifications of conceptual meanings can also be viewed from a more linguistic point of view. For example, it is reasonable to expect that it is not available for the meaning of light verbs (*have, be, make*), morphological categories (plural, singular, tense, mood), thematic roles, and syncategorematic expressions (*and, or, if-then*).[10] Moreover, there is evidence that it depends on the syntactic category of expressions to what extent it is applicable. For example, there is a striking difference between the meaning of the verb *exist* and the meaning of the nominalization *existence*. Many philosophers, such as van Inwagen (1998), take existence to be a notion that trivially applies to everything there is, or at least every actual thing. Their use of the noun *existence* would be perfectly legitimate for talking about that notion. However, in natural language, the predicate *exist* is subject to strict conditions on the type of entity to which it can apply, it applies to material and abstract objects, but not to events (Hacker 1982; Cresswell 1986; Moltmann 2013c, 2020):

(1) a. The house still exists.
    b. The largest prime number does not exist.

(2) a. ??? The rain still exists.
    b. ??? The protest existed yesterday.

Events do not 'exist', but 'take place', 'happen', 'occur', or 'last'. The reason why *exist* exhibits such restrictions, arguably, is that *exist* has primarily a time–relative meaning and conveys the complete presence of an entity throughout a time (Moltmann 2013c, 2020). By contrast, the verb *existence* can be used without imposing any restrictions on the entities it applies to (as in *the existence of everything there is*). What is important is that the verb *exist* cannot be used so as to convey the unrestricted notion of existence. The resistance of *exist* to events is robust: even a philosopher convinced of the univocality of the concept of existence (such as van Inwagen) will be unable to apply *exist* to events.[11] This appears to be a

[10]In generative syntax, this should generally hold for what is called functional categories, as opposed to lexical categories.

[11]There also are various sorts of philosophers that have particular views about existence, for example that only fundamental or mereologically primitive entities exist. As long as the sortal restrictions of *exist* are respected, such views can attach to the use of the verb as well.

reflection of the semantic core-periphery distinction: the meaning of the verb *exist* belongs to the semantic core of language, whereas at least one of the meanings of the noun *existence* belongs to the semantic periphery. It is also a reflection of the ontological core-periphery distinction: whenever an expression conveying a metaphysical notion undergoes legitimate conceptual modification, this involves a shift from the semantic as well as the ontological core of language to its semantic as well as its ontological periphery.[12]

The semantic core-periphery distinction is very similar to the ontological core-periphery distinction. The semantic core involves concepts that are implicit in the ordinary use of language, that is, that speakers implicitly accept when they use language in the ordinary way. The semantic periphery involves doing something to ordinary conceptual meanings ('conceptual engineering') and thus adding to the given domain of lexical meaning. The ontological periphery was characterized as involving ontological reflection, and thus adding to the ontological domain that a speaker accepts when using the language. The close connection follows from the fact that metaphysical concepts also belong to the ontological domain, and conceptual engineering would send them from the ontological core to the ontological periphery.[13]

Another feature that the semantic and ontological core-periphery distinctions share besides a distinction between implicit acceptance and reflection is a form of true creativity that language, with its conceptual and ontological domain, permits, with choices to be made in response to particular demands or interests, quite unlike the creativity Chomsky attributed to language, which involves applying a given set of rules to produce utterances of new sentences or expressions. Creativity in the latter sense is a general feature of the morphosyntax of language, allowing for the production of an infinite number of new sentences and words. But this is not the sort of expansion that leads to a periphery. The conceptual and ontological structures associated with language allow for creativity that goes along with distinctive cognitive acts, leading to a periphery in the relevant sense.

[12] Somewhat similar observations can be made about the adjective *true* and the nominalization *truth*. Philosophers have various different views about truth and can use the nominalization *truth* to convey their respective concept of truth. Philosophers (and perhaps non-philosophers) also generally have the view that representational mental objects (with a mind-world direction of it) are truth bearers. However, *true* fails to apply to some of the objects that one would think are truth bearers, for example impressions and speculations (an impression cannot be said to be true or false, and neither can a speculation) (Moltmann 2018). Of course, there may be various reasons (not just conceptual fixed points) why *true* exhibits additional restrictions in relation to the nominalization *truth*. But in any case, *truth* conveys a reflective notion that is not the same as that conveyed by *true*.

[13] Moreover, the switch from the adjective to the noun goes along with reification of a concept (adjective) as an object (noun). So here it is not just concept modification, but also reification, and thus an ontological operation.

13.4.3 The Core-Periphery Distinction in Syntax and Phonology

Chomsky (1981, 1986, 1998) introduced a core-periphery distinction for syntax. Very roughly, those parts of a natural language are considered in the periphery that are anomalous or 'added on' from influences from other languages, whereas the core of language reflects universal grammar.[14] Chomsky's distinction at first sight does not seem co-extensive with the semantic and ontological core-periphery distinctions. Various sorts of complex or derived expressions that would be part of the conceptual or ontological periphery should belong to the syntactic core, including nominalizations (e.g. *existence*) and reifying terms. However, the syntactic core-periphery distinction arguably goes along with the distinction between functional categories and the lexicon (Yang 2016), which means that the peripherical status of nominalizations and reifying terms could be attributed to their involvement of full nouns (sortals in the case of reifying terms). Thus there may be a sense in which the syntactic and the ontological periphery coincide after all.

There are other features that the two sorts of peripheries share. One intuitive feature is that of being "additional" with respect to the core. The ontological and semantic peripheries involve additional cognitive acts besides implicit acceptance of the domain of the core: acts of philosophical reflection of some sort. In addition, the two sorts of peripheries have the purpose of separating universal features of linguistic or cognitive systems from those that are not (whether due to outside influence or explicit reflection). Only when focusing on the core of language is the pursuit of universals in the cognitive ontology reflected in language possible, such as the Abstract Objects Hypothesis.

There is another core-periphery distinction that has been used in generative grammar, namely in phonology. According to that distinction, the lexicon has a core-periphery organization according to degrees of assimilation/integration of the vocabulary (Îto and Mester 1995a, b). Here the core-periphery distinction is not a binary distinction, but a gradual one, with different types of items being more or less in the periphery or core. The phonological core-periphery distinction is also based on the feature of natural language being expandable, but now allowing for the vocabulary to be expanded by importing items from other languages. The semantic and ontological peripheries, by contrast, involve expansion from within the conceptual and ontological system.

[14] Yang (2016) contains a more recent discussion and defense of the core-periphery distinction in syntax.

13.5 Abstract Objects and the Core-Periphery Distinction: A Potential Universal of Natural Language Ontology

With the distinction between core and periphery in the ontological domain of natural language we can turn to the Abstract Objects Hypothesis, given again below:

(3) The Abstract Objects Hypothesis

Natural language does not involve reference to abstract objects in its ontological core, but only in its ontological periphery.

This hypothesis is based on a range of more specific generalizations regarding various expressions in the core of English that have been taken to involve reference to abstract objects (Moltmann 2013a). The various expressions in question are reanalyzed in one of the following four ways:

[1] The expression involves no reference to an abstract object, but instead to a concrete one, for example a trope (particularized property).
[2] The expression involves no reference to a single abstract object, but plural reference to various actual or possible particulars.
[3] The expression does not involve reference to a truly abstract object, but rather reference to an object that strictly inherits all its properties from actual or possible concrete entities.
[4] The expression does not act as a referential term (in relevant environments), but rather as a nonreferential complement or subject.

Analysis [3] avoids a particular notion of an abstract object, as an object that bears properties directly. When an object strictly inherits all its properties from concrete entities, then the truth conditions of statements about the object generally reduce to those of a statements just about those concrete objects.

In what follows, I will just briefly indicate how various putative abstract terms are reanalyzed, referring the reader to Moltmann (2013a) for empirical and formal semantic details.

First, bare (determinerless) adjective nominalizations such as *wisdom* have standardly been considered singular terms standing for properties (Hale 1983). Philosophers generally make use of such terms and not explicit property-referring terms such as *the property of being wise* when arguing for natural language involving properties as objects. That is because *wisdom* is regarded a term in the core of language and *the property of being wise* a term in the periphery. The problem for the standard view is that *wisdom* displays different readings with various sorts of predicates from those displayed by its explicit property-referring counterpart. This is illustrated below (Moltmann 2013a):

(4) a. Wisdom exists.
 b. The property of wisdom exists.
(5) a. John found wisdom.
 b. ??? John found the property of wisdom.
(6) a. Wisdom is admirable.
 b. ??? The property of wisdom is admirable.
(7) a. True wisdom is rare.
 b. ??? The property of being truly wise is rare.

(4a) can state only the existence of an instance of wisdom, not the existence of an abstract object as such, unlike (4b). (5a) means that John found an instance of wisdom, not an abstract property object, unlike (5b) (which could be true only in a metaphysical fantasy). (6a) means that instances of wisdom are admirable, not the abstract property as such, unlike (6b). Predicates like *rare*, which only care about the distribution of instances, sound natural only with bare adjective nominalizations, as in (7a), not with explicit property-referring terms as in (7b). Generally, predicates are true of what bare adjective nominalizations stand for in virtue of properties of particular instances; they do not attribute a property to an abstract object, unlike with explicit property-referring terms. In Moltmann (2004), I had adopted analysis [3] positing a type of entity (an Aristotelian 'kind') which is unable to bear properties directly but only by inheritance from its instances, namely a kind whose instances are particular tropes. In Moltmann (2013a), bare adjective nominalizations like *wisdom* are no longer considered singular terms, but are taken to stand for kinds conceived as 'modalized pluralities', pluralities (as many) of all the actual and possible tropes (particularized properties) (that is, they refer plurally in the sense of Yi 2005, 2006). In both cases, a predicate P when applying to a kind (an Aristotelian kind or modalized plurality) has a derivative meaning, being true of the kind just in case the original property expressed by P is true of some instances (*exist* and *find*) or instances in general (*admirable*). These accounts extend to bare plurals (*giraffes, blue pens*) and bare mass nouns (*water, white rice*), which again are not considered singular terms standing for kinds as abstract objects.

Entities that strictly inherit all their properties from concrete ones are also posited as semantic values of definite noun phrases of the sort *the water in the container* (with predicates like *decrease*) and *the book John needs to write* (see also Moltmann to appear). These are variable objects, entities which at different times and in different situations have possibly different manifestations, from which they inherit their properties.

In natural language semantics, it has become standard to make use of abstract objects that are degrees, namely for the analysis of positive and comparative adjectives. *Tall* (in *John is tall*) means taller than a contextually given standard degree and *taller* expresses a relation between two pairs each consisting of an individual and a degree. The apparent involvement of degrees in the semantics of adjectives is reanalyzed in terms of [1] and [2], replacing abstract degrees by tropes or kinds of tropes (Aristotelian kinds or modalized pluralities) (see also Moltmann

2009). On that view, *John is taller than Mary* means 'John's height (a quantitative trope) exceeds Mary's height), and *John is tall* 'John's height exceeds the kind of tropes that makes up the contextual standard'.

Numbers as abstract objects have generally been taken to be well-reflected in natural language. Frege, in particular, appealed to natural language when arguing for numbers being objects. For Frege, terms like *the number of planets* as well as simple number words like *eight* are singular terms standing for objects. Such apparent number-referring terms are reanalyzed adopting analyses [1] and [4] (see also Moltmann 2013d). First, number words like *eight* are no longer considered referential terms when occurring in argument position, but rather expressions that retain the meaning they have as noun modifiers (see also Hofweber 2005). The meaning of arithmetical statements in natural language then involves what in the philosophy of mathematics is called 'the Adjectival Strategy' (Dummett 1973, Hodes 1984). Roughly, *Two and two is four* on that view is analysed as 'if there were two things and another two thing, then there would be four things'. Apparent number-referring terms like *the number of planets* are reanalyzed as terms referring to number tropes (the planets reduced to just how many they are).

Propositions are abstract objects that play a central role in the semantics of attitude reports and other sentence-embedding constructions. Propositions are considered problematic not just because of their abstractness, but because of specific problems of their own, their role as truth bearers and contents of attitudes (the problems of the truth-directedness and of the unity of the proposition).[15] At least since Frege, propositions have been taken to be the semantic values of *that*-clauses (and sentences in general) as well as the contents of attitudes. Propositions have to be abstract, so the Fregean view, since contents of attitudes are sharable among different agents. The apparent compositional semantics of attitude reports appears to require propositions, on the assumption that *that*-clauses act as arguments providing an argument of the relation expressed by the embedding verb. However, there is good evidence that *that*-causes do not in fact act as referential terms and thus that their meaning does not have the status of an object. The linguistic evidence includes the failure of *that*-clauses to be substitutable by *the proposition that* S with most attitude verbs. In Moltmann (2013a), I proposed that *that*-clauses instead act as plural terms standing for an ordered plurality of propositional constituents. More recently, I adopted the view that they act semantically as predicates of an object associated with the attitude verb (what I call an 'attitudinal object') (Moltmann 2014). Special quantifiers like *something* or *everything* on the traditional view have been taken to act as propositional quantifiers. But again there is strong linguistic evidence that they do not behave that way, but rather act as nominalizing quantifiers ranging over attitudinal objects or kinds (pluralities) of them, the sorts of thing nominalizations of attitude verbs stand for. The view extends to other sentence-embedding constructions such as truth predicates and modals (Moltmann 2018). Given this view, propositions no longer play a role in the semantics of natural

[15] See Moltmann (2013a, Chapter 5) and references therein.

language. They at best play a role as semantic values of explicit proposition-referring terms, such as *the proposition that* S.

Expression types are also abstract objects that are generally held to be objects of reference in natural language, namely referents of pure quotations. But like *that*-clauses, pure quotations display features of nonreferentiality, for example by not generally permitting substitution by an explicit expression-referring term (*'Red' means 'rouge'* does not imply *'Red' means the expression 'red'*, which is unacceptable). Instead of taking pure quotations to be referential terms standing for expression types, their function is now considered that of 'presenting themselves', forming a complex predicate with the embedding verb.

There are natural language terms explicitly referring to propositions, properties, numbers degrees, and expression types, of the sort *the number eight*, *the proposition that* S, *the word 'rouge'*, *the sentence* S. But those terms belong to the ontological periphery of language, not its core. When it comes to the core of natural language, what appeared to be expressions referring to abstract objects are now considered either expressions referring to concrete particulars or kinds of them or expressions that do not act as referential terms in the first place (number words, *that*-clauses, quotations), but contribute to the meaning of the sentence in a different way.

The ontology of the core of language, according to the Abstract Objects Hypothesis, thus is an Aristotelian ontology of just concrete entities, or at least objects whose involvement in a statement would guarantee truthconditional equivalence with a statement just about concrete entities.

13.6 The Ontology Reflected in Abstract Terms in the Periphery

On the present view, expressions and uses of expressions in the ontological periphery are legitimate parts of language (or its use). The question then arises for the theoretician how to handle their semantics and ontology.

Of course, the semanticist cannot but accept the meanings that expressions are meant to obtain on a non-ordinary use, and so for newly introduced 'technical' expressions. But there is a question of how to regard the ontology that goes along with referential noun phrases on a non-ordinary or technical use. Here it is important to keep in mind that the ontology of natural language in general should be understood as an ontology of appearances. A referential noun phrase that is part of the ontological periphery thus should have a semantic value that is an object, but which may be a merely conceived object (though it may also turn out to be an actual one).

There are also referential noun phrases whose semantic values are part of the periphery due to their compositional semantics. In particular, these would be reifying terms of the sort *the number eight*, *the proposition that* S, and *the property of being wise*. While this is not the place to discuss the syntax and semantics of reifying terms in detail, one general condition that should certainly obtain is that their semantics needs to be sufficiently general to allow for them to have semantic values on any given philosophers' theory of abstract objects. Yet the compositional semantics of reifying terms is also indicative of the role and nature of abstract objects in the core of natural language. The structure of reifying terms arguably involves a nonreferential expression or use of an expression following the sortal: *eight* in *the number eight* is still an adjective, and as such just mentioned rather than used (Moltmann 2013a). This then suggests a context-dependent semantics of reifying terms along the following lines. In a context of use c, the reifying term *the* N X stands the object *o* that is obtained on the basis of statements in which X occurs (referentially or nonreferentially), where it depends on the background assumptions in c what S consists in. If the object o is conceived as an object whose nature is exhausted by the attribution of predicates obtained from true statements in which X occurs, this would amount to a pleonastic account of abstract objects (Schiffer 1996, 2003). For example, given the Adjectival Strategy, *divisible by two* in *eight is divisible by two* roughly means 'any possible plurality of eight things consists in two equal-membered subpluralities'. Then the predicate *is divisible by two* can be attributed to o in virtue of the pleonastic equivalence below:

(8) *The number eight* is divisible by two iff eight is divisible by two.

Similarly, explicit property-referring terms would introduce objects whose nature is exhausted by pleonastic equivalences such as the one below (Schiffer 2003):

(9) John has *the property of being happy* iff John is happy.

However, the semantics of reifying terms should not commit itself to a pleonastic account of abstract objects, that is, an account on which there is nothing more to a number or a property than what can be attributed to it in virtue of an equivalence such as (8) or (9). Rather it should be compatible with various other philosophical views of abstract objects and be able to accommodate the use of sortals (in reifying terms) based on philosophical definition. Thus, (8) is compatible with a view on which *the number eight* is a set-theoretical construct or else a type of collection or just a light object whose properties are to be read of equivalences as in (8). Similarly, (9) is compatible with a view on which properties are platonic objects, collections of similar tropes, or else light objects, 'mere shadows of predicates'. The pleonastic account of abstract objects as semantic values of sortal nouns is no longer appropriate to pursue once the ontological periphery of natural language is acknowledged with its space for multiple ontologies that serve the semantic values of sortal nouns.

What natural language tells us about abstract objects is not that they could not be objects of reference (and of the sort that a particular philosophical theory takes them to be). It only tells us that they are not part of the ontological core of natural language. They can be part of the ontological periphery, and (at least as merely conceived objects) fall under various philosophical views, something that semantic theory needs to be able to accommodate.

13.7 Conclusion

The core-periphery distinction is crucial for natural language ontology, and it is particularly important when approaching abstract objects from the point of view of natural language. It is the basis for the Abstract Objects Hypothesis, a putative universal of natural language ontology.

The ontological core-periphery distinction is due to the legitimate expandability of language, the same feature of language that is grounds for a distinction between core and periphery for lexical/conceptual meaning and that permits 'conceptual engineering'. In both cases, expansion goes along with an additional cognitive effort of reflection. In the case of the ontological periphery, such an act of reflection consists in the recognition of an object as belonging to a particular ontological category, by the application of a sortal noun. Recognizing both core and periphery allows taking language with its ontology to be a cognitive system that comprises both implicit acceptance and reflection and gives justice both to the ontology implicit in natural language and the reflective ontology of philosophers and non-philosophers.

Acknowledgments Part of the material of this paper has been presented in courses on natural language ontology in Duesseldorf, ESSLLI in Sofia (Bulgaria), and Munich in 2018 as well as at NYU (May 2019) and the IHPST, Paris (July 2019). The paper has benefitted greatly from discussions with the audiences as well as from numerous conversations with Kit Fine, exchanges with Noam Chomsky and Donca Steriade, as well as comments by Matti Eklund, Jonathan Schaffer, and an anonymous referee on a previous version of this paper.

Appendix: Some Potential Issues for the Abstract Object Hypothesis

In this appendix, I will briefly discuss some issues that appear problematic for the Abstract Objects Hypothesis.

First, there is a notion of an abstract state that, it has been argued, plays a role not just for the semantic values of reifying terms of the sort *the state of being happy*, but also as implicit arguments of stative verbs. Thus, Maienborn (2007) introduced a distinction between abstract states ('Kimian states' as she calls them Kim 1976) and concrete states ('Davidsonian states' as she calls them Davidson 1967). Both sorts of states, she argues, act as arguments of stative verbs. Abstract states are implicit arguments of of abstract state verbs such as *owe, own, know, be, have*, and *resemble* (see also Moltmann 2011). Unlike concrete state verbs like *sleep* and *sit*, abstract state verbs do not accept a range of adverbial modifiers, such as causal and locational adverbials, and that is, so Maienborn, because of the ontology of abstract states, which lack causal roles and a spatial location. Abstract state verbs certainly are part of the core of language, which is in conflict with the Abstract Objects Hypothesis. Maienborn's view, though, is not uncontroversial. There are alternative accounts of the ‘stative adverb gap’, as it is called. One of them takes stative verbs to lack an event argument position to be filled in by states (Katz 2003). Another account takes adverbial modifiers to apply to truth makers rather than event arguments of verbs, or at least when modifying stative verbs (Moltmann 2007).

Another apparent issue for the Abstract Objects Hypothesis is kinds and the applicability of existence predicates (Moltmann 2020). Existence predicates impose particular conditions on what sorts of objects they can apply to, with or without temporal or spatial modifier. Roughly, the generalization is that existence predicates can apply to an entity o relative to a location l just in case o is completely present throughout l (Moltmann 2020). The complete presence condition throughout a location has the consequence that existence predicates with temporal modifiers apply only to enduring objects and not events (Sect. 13.4.2). With spatial modifiers existence predicates can apply to only few sorts of entities (illnesses, languages, and kinds), and that is because those sorts of objects have abstract part structures permitting them to be present at multiple spatial locations at once. As entities with an abstract part structure, kinds then will have to be abstract objects themselves (rather than, say, pluralities of possible and actual particulars). However, other predicates than existence predicates just do not treat kinds as abstract objects, as we have seen in Sect. 13.5. There is a solution to this puzzle, and that is to take the constitutive features of kinds not to be abstract properties, but rather kinds of tropes (particularized properties). Kinds of tropes are present throughout a location just in case they have instances throughout the location.

Another issue with the core-periphery distinction is that philosophers drawing on natural in support of an ontological view sometimes make use of expressions from the ontological periphery of language. For example, Vendler (1967) made use of the reifying term *the fact that* S to argue for a distinction between facts and events. I myself in Moltmann (2004, 2013a) made of use of reifying terms like *the property of being wise* to argue for a distinction between properties and kinds of tropes (the semantic values of terms like *wisdom*). Here the use of reifying terms is legitimate as it serves to show an ontological difference between a type of entity from the core and a type of entity already in the periphery. The term from the periphery is associated with intuitions ordinary speakers have – the periphery is not reserved for philosophers only. Of course, using technical terms that a particular philosopher may have introduced in the context of a philosophical theory would be a different matter.

References

Cappelen, H. (2018). *Fixing language: An essay on conceptual engineering*. Oxford: Oxford University Press.

Carlson, G. (1977). A unified analysis of the English bare plural. *Linguistics and Philosophy*, 413–457.

Chalmers, D. (2011). Verbal disputes. *Philosophical Review, 120*(4), 515–566.

Chomsky, N. (1981). *Lectures on government and binding. The Pisa lectures* (Studies in Generative Grammar, no. 9). Dordrecht: Foris Publications.

Chomsky, N. (1986). *Knowledge of language. Its nature, origin, and use*. Westport/London: Praeger.

Chomsky, N. (1998). *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.

Cresswell, M. J. (1986). Why object exists, but events occur. *Studia Logica, 45*, 371–5.

Davidson, D. (1967). The logical form of action sentences. In N. Rescher (ed.). *The Logic of Decision and Action* (pp. 81–95). Pittsburgh: Pittsburgh University Press.

Dummett, M. (1973). *Frege: Philosophy of language*. Cambridge, MA: Harvard University Press.

Eklund, M. (2015). Intuitions, conceptual engineering, and conceptual fixed points. In C. Daly (Ed.), *Palgrave handbook of philosophical methods*. New York: Palgrave Macmillan.

Fine, K. (1991). The study of ontology. *Nous, 25*(3), 262–294.

Fine, K. (2017). Naïve metaphysics (J. Schaffer, Ed.). *Philosophical Issues*, *27*.

Frege, G. (1884). *Die Grundlagen der Arithmetik* (Foundations of Arithmetic).

Frege, G. (1918/9). Thoughts. In B. McGuinness (ed.). *Collected Papers on Mathematics, Logic, and Philosophy* (pp. 351–372). Blackwell, Oxford, 1984.

Hacker, P. M. S. (1982). Events, Ontology, and Grammar. *Philosophy 57*, 477–486.

Hale, B. (1987). *Abstract Objects*. Blackwell, Oxford.

Hodes, H. (1984). The ontological commitment of arithmetics. *Journal of Philosophy, 81*, 123–149.

Hofweber, T. (2005). Number determiners, numbers, and arithmetics. *Philosophical Review, 114*(2), 179–225.

Itô, J., & Mester, A. (1995a). Japanese phonology. In J. Goldsmith (ed.). *The handbook of phonological theory* (pp. 816–838). Malden: Wiley Blackwell.

Itô, J., & Mester, A. (1995b). The core-periphery structure of the lexicon and constraints on Reranking. In J. Beckman, S. Urbanczyk, & L. Walsh (Eds.), *University of Massachusetts Occasional Papers in Linguistics [UMOP]* (Vol. 18, pp. 181–209). Amherst: University of Massachusetts/GLSA.

Katz, G. (2003). Event arguments, adverb selection, and the stative adverb gap. In E. Lang (Eds.). *Modifying Adjuncts*, de Gruyter.

Kim, J. (1976). Events as Property Exemplifications. In M. Brand & D. Walton (Eds.). *Action Theory*. Reidel: Dordrecht, pp. 310–326.

Maienborn, C. (2007). On Davidsonian and Kimian states. In I. Comorovski / K. von Heusinger (eds.): Existence: Semantics and syntax. Springer, Dordrecht, 107–130.

Moltmann, F. (2004). Properties and kinds of tropes. New linguistic facts and old philosophical insights. *Mind, 123*(1), 1–41.

Moltmann, F. (2007). Events, tropes, and truthmaking. *Philosophical Studies, 134*, 363–403.

Moltmann, F. (2009). Degree structure as trope structure A trope-based analysis of comparative and positive adjectives. *Linguistics and Philosophy, 32*, 51–94.

Moltmann, F. (2013a). *Abstract objects and the semantics of natural language*. Oxford: Oxford University Press.

Moltmann, F. (2013b). On the distinction between abstract states, concrete states, and tropes. In A. Mari, C. Beyssade, & F. Del Prete (Eds.), *Genericity* (pp. 292–311). Oxford: Oxford University Press.

Moltmann, F. (2013c). The semantics of existence. *Linguistics and Philosophy, 36*(1), 31–63.

Moltmann, F. (2013d). Reference to numbers in natural language. *Philosophical Studies, 162*(3), 499–534.

Moltmann, F. (2014). Propositions, attitudinal objects, and the distinction between actions and products. *Canadian Journal of Philosophy, 43*(5–6), 679–701.

Moltmann, F. (2017). Natural language ontology. *Oxford Encyclopedia of Linguistics,* online.

Moltmann, F. (2018). Truth Predicates, Truth Bearers, and their Variants. In J. Wyatt (ed.). *Special issue Truth: Concept meets Property, Synthese*, online first.

Moltmann, F. (2020). Existence Predicates. *Synthese, 197*(1) 311–335

Moltmann, F. (2019). Natural language and its ontology. In A. Goldman & B. McLaughlin (Eds.), *Metaphysics and cognitive science* (pp. 206–232). Oxford: Oxford University Press.
Moltmann, F. (to appear). Variable objects and Truthmaking. In M. Dumitru (Ed.), *The philosophy of Kit Fine*. New York: Oxford University Press.
Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, MA: MIT Press.
Schiffer, S. (1996). Language-created and language-independent entities. *Philosophical Topics, 24*(1), 149–167.
Schiffer, S. (2003). *The things we mean*. Oxford: Clarendon Press.
Strawson, P. (1959). *Individuals. An essay in descriptive metaphysics*. London: Methuen.
van Inwagen, P. (1998). Meta-ontology. *Erkenntnis 48*, 233–250.
Vendler, Z. (1967). *Linguistics in philosophy*. Ithaca: Cornell University Press.
Williams, D. (1953). On the elements of being. *Review of Metaphysics, 7*, 3–18.
Wright, C. (1983). *Frege's Conception of Numbers as Objects*. Aberdeen UP: Aberdeen.
Yang, C. (2016). *The price of linguistic productivity. How children lean to break the rules of language*. Cambridge, MA: MIT Press.
Yi, B.-Y. (2005). The logic and meaning of plurals. Part I. *Journal of Philosophical Logic, 34*, 459–506.
Yi, B.-Y. (2006). The logic and meaning of plurals. Part II. *Journal of Philosophical Logic, 35*, 239.

Chapter 14
How to Vindicate (Fictional) Creationism

Alberto Voltolini

Abstract In this paper, I want to show that Moderate Creationism (MC), a variant among realist doctrines concerning fictional entities (*ficta*), is the only form of Fictional Creationism that may rescue it from both the ontological and the metaphysical criticisms that have been recently raised against this position in general. According to MC, what is distinctively required in order for a *fictum* to come into existence *qua* abstract mind-dependent entity is that a certain *reflexive stance* applies to the non- ontologically committal make-believe practice that there is a certain (nonactual yet typically concrete) individual. In this stance, the practice is taken as involving a certain set of properties, the properties the storyteller mobilises in the relevant bit of her narration. So taken, the practice amounts to the fact that that very set of properties is make-believedly identical with a certain (nonactual yet typically concrete) individual. In particular, I appeal to a strong version of MC, according to which, once it is so taken, the practice *counts* as a certain *fictum*, in virtue of a constitutive rule for its generation. For this version may account for the special sense in which *ficta* may be created entities.

Keywords Fictional Creationism · Radical Creationism · Moderate Creationism · Ontology · Metaphysics

14.1 Introduction

In this paper, I want to show that Moderate Creationism (MC), a variant among realist doctrines concerning fictional entities (*ficta*), is the only form of Fictional Creationism that may rescue it from both the ontological and the metaphysical criticisms that have been recently raised against this position in general. According to MC, what is distinctively required in order for a *fictum* to come into existence

A. Voltolini (✉)
Department of Philosophy and Education Sciences, University of Turin, Turin, Italy
e-mail: alberto.voltolini@unito.it

J. L. Falguera, C. Martínez-Vidal (eds.), *Abstract Objects*, Synthese Library 422,
https://doi.org/10.1007/978-3-030-38242-1_14

qua abstract mind-dependent entity is that a certain *reflexive stance* applies to the non- ontologically committal make-believe practice that there is a certain (nonactual yet typically concrete) individual. In this stance, the practice is taken as involving a certain set of properties, the properties the storyteller mobilises in the relevant bit of her narration. So taken, the practice amounts to the fact that that very set of properties is make-believedly identical with a certain (nonactual yet typically concrete) individual. In particular, I appeal to a strong version of MC, according to which, once it is so taken, the practice *counts* as a certain *fictum*, in virtue of a constitutive rule for its generation. For this version may account for the special sense in which *ficta* may be created entities.

14.2 Fictional Creationism and Its Varieties

Fictional Creationism (from now on, simply *Creationism*) is an ontologico-metaphysical position constituted by two theses. The first is the *ontological* thesis (a): in the overall ontological inventory, there also are fictional entities (*ficta*). The second is the *metaphysical* thesis (b): *ficta* depend for their existence on the ultimately imaginative activities of certain people – basically, make-believe practices of certain sorts, primarily storytelling practices – which make them abstract entities of a certain kind, mind-dependent *abstracta*.[1]

To start with, Creationism shares (a) with any other realist-oriented position about *ficta*; namely, any position holding that, in the overall ontological inventory, there also are *ficta*: e.g. (Neo)Meinongianism, (Im)Possibilism, and their combinations.[2] Thus, (b) is the qualifying thesis of Creationism, for it distinguishes it metaphysically from the above positions. Furthermore, Creationism itself comes in different metaphysical varieties, depending on how thesis b) is further modulated.

Radical Creationism (RC) holds that a *fictum*'s existence supervenes on the existence of a certain imaginative practice. RC can be further interpreted either *weakly* or *strongly*, depending on whether the supervenience claim points to a mere *factual* form of supervenience (Searle 1979; Salmon 1987, 1998) or rather to a *conceptual* form of supervenience (Schiffer 1996, 2003; Thomasson 2003a, b). According to such variants, a conditional of the kind "If one mobilises the name 'NN' within a make-believe storytelling practice, then 'NN' refers to a fictional entity" is taken to be either a merely actual or rather a conceptual truth.

[1] Thesis b) is mostly explicitly defended in Thomasson (1999), but it traces back to the seminal work of Ingarden (1973) as well as to van Inwagen (1977) and Kripke (2013). For other defenders of the thesis cf. Searle (1979), Salmon (1998), Schiffer (1996, 2003), Predelli (2002), Goodman (2004), Braun (2005).

[2] As to (Neo)Meinongians, cf. e.g. Castañeda (1989), Parsons (1980), Zalta (1983). As to (Im)possibliists, cf. Lewis (1983), Priest (2016)[2], Berto (2013). Berto's and Priest's theories are also a variant of (Neo)Meinongianism, i.e., Modal Meinongianism.

By contrast, *Moderate* Creationism (MC) holds that such a make-believe practice is just a necessary, but not a sufficient condition for a *fictum* to come into existence. What is further required for that existence is that a certain *reflexive stance* on that make-believe practice takes place. In this stance, the practice is taken as involving a certain set of properties, typically the properties a storyteller mobilises in the relevant bit of her narration. So taken, the practice is no longer the non- ontologically committal practice that there is a certain nonactual yet typically concrete individual; instead, it amounts to the fact that that very set of properties is made believe to be identical with that individual. In Evans' (1982) terminology, the practice is no longer the *creative* make-believe game that there is a certain nonactual yet typically concrete individual, it is a *conservative* make-believe game concerning a set. In this way of taking the practice, *no fictional individual* is mobilised yet: to repeat, in this way the practice amounts to the fact that, *of* a certain property set, one makes believe that it is identical with a typically concrete individual. For example, in the Sherlock Holmes tales, one makes believe of a certain property set {*being a detective*, *being clever*, *being cocaine-addicted* ...} that it is identical with a concrete human being. Once that practice is so taken, a *fictum* comes into being as the abstract mind-dependent entity it is; so taken, the practice that there is a nonactual yet typically concrete individual *counts* as *this other*, abstract, individual. The reflexive stance is *public*, insofar as its making the practice count as a *fictum* must display a *normative* force: the *fictum* that arises out of the stance is the *right* fictional entity that complies with the constitutive rule for its generation that stance has put in force.[3]

Various people – both realists and antirealists on *ficta* – think that Creationism is the best realist account of fictional entities.[4] Yet first of all, MC comes on stage because RC, in both its variants, suffers from various problems (Voltolini 2009). On the one hand, the supervenience thesis characterizing RC hardly amounts to a conceptual truth. Consider the antirealist fictionalist position on *ficta* tracing back to Evans (1982) and Walton (1990). This position holds that entertaining certain

[3]Voltolini (2006, 2015b). One normally believes that Kripke (2011, 2013) is a defender of RC. Yet there are reasons to call this belief into question. First, Kripke says that the make-believe practice in which one pretends that there is an individual that does such and such things is not ontologically committed to *ficta* (2013: 81, 148); *ficta* rather come into existence in virtue of certain human activities. Kripke gives, as examples of such activities, telling stories, writing plays, writing novels (2011:63–4), (2013: 73, 76). Second, he says that fictional works are sufficient conditions for *ficta* (2013: 71–2). Thus, one may suspect that within such activities there also is the sort of reflection that is manifested in what he labels a derivative or extended use of language that is committed to *ficta* (*ibid.*: 81, 103, 148–50). Schiffer (1996, 2003) called this use the *hypostatising* use, which is for Kripke the condition for further ascribing *ficta* also story-relative properties (2011: 65), (2013: 74–5, 83). All this suggests that Kripke is rather a forerunner of MC.

[4]For Sainsbury (2010: 61–3, 82–5), who is an antirealist, Creationism is the only realist doctrine that manages to face the socalled *selection problem*: how is it that an author manages to select one rather than another *fictum* candidate among countless candidates? For Zvolenszky (2015a), who is a realist, Creationism is the only realist doctrine that manages to face the epistemological problem Kripke (1980) raised consisting in finding both a historical connection and a suitable mode of introduction for a name to refer to a *fictum*.

make-believe practices determines no ontological committment to *ficta*. Not only make-believe practices, are clearly noncommitted to *ficta*: in their own scope and within that scope only, there is just a committment to nonactual yet typically concrete individuals. But also, they go on saying, such practices are able to account for whatever seemingly involves *ficta*. Now, if the supervenience thesis amounted to a conceptual truth, the fictionalist position would, absurdly, be unintelligible; as if a fictionalist had no mastery of the very notion of a *fictum*.[5] On the other hand, the supervenience claim does not even amount to a merely actual truth.[6] For, as I just recalled and even creationists admit, *outside* the scope of the make-believe, a make-believe practice is not ontologically committed to the items it posits *within* that scope: making believe *that* there is a certain (nonactual yet typically concrete) individual does not entail that there is such an individual in the overall ontological inventory. A fortiori, the practice is not ontologically committed to the entities of a different metaphysical kind *ficta* amount to *qua* abstract mind-dependent things. By itself, like any other non-committal practice, a make-believe practice ontologically leads nowhere else.[7]

Moreover and more importantly, MC is really required in order for someone to be a creationist. For, as I will try to show in Sects. 14.3 and 14.4 respectively, it seems to be the only way to bypass some further powerful criticisms that have been recently raised against Creationism in general: the *ontological* criticisms (Kroon 2011, 2013, 2015) and the *metaphysical* criticisms (Deutsch 1991; Brock 2010). In the proceeding of this paper, I will indeed try to show why, in order to dispense with such criticisms, a creationist must stick to MC, taken moreover in its strong variant that I have already implicitly sketched above. According to strong MC, there is a *constitutive*, not causal, relation between the reflexive stance and the *fictum* that it generates.

14.3 The Ontological Criticisms

Kroon's arguments intend to dismantle a famous semantic-ontological argument that realists about *ficta* have often appealed to. The argument was originally defended by van Inwagen, who is generally considered to be a forerunner of creationism. Notoriously, the Van Inwagen argument (VA) runs as follows:

---

5 Voltolini (2006). For a more radical critique on this concern see Everett (2013:131–2), who holds instead (wrongly, I think: cf. fn.28) that our notion of a *fictum* involves that such an entity does not exist, thus supporting an antirealist intuition on these matters.

6 Cf. Yagisawa (2001).

7 Curiously enough, Meinong himself seemingly was of this idea in the first 1902 edition of his (1983). At the time of the second 1910 edition, however, he had probably changed his mind, by holding a supervenience claim that makes an assuming (or make-believe) practice to be ontologically generative. Cf. Kroon (1992).

(a) If (once suitably regimented in a first-order language expressing scientific statements) some existentially quantified sentences from literary criticism seemingly involving *ficta* are genuinely true, then such sentences ontologically commit us to such entities;
(b) Yet some such existentially quantified sentences seemingly involving *ficta* are genuinely true;
(c) Hence, such sentences ontologically commit us to such entities.[8]

Suppose one takes literary criticism as a scientific theory to be regimented in first-order language. In that criticism, says van Inwagen, there surely are genuinely true existential sentences that provide an ontological committment to *ficta*, e.g.:

(1) There are characters in some 19th-century novels who are presented with a greater wealth of physical detail than is any character in any 18th-century novel.
(2) Some characters in novels are closely modelled on actual people, while others are wholly products of the literary imagination, and it is usually impossible to tell which characters fall into which of these categories by textual analysis alone.

Kroon's counterstrategy consists in providing a two-step argument that puts premises (a)–(b) of VA into question, by supplementing a provisional rejection of (a) with the effective rejection of (b). *Contra* (a), even if there were true existential sentences seemingly involving *ficta*, we would not be forced to accept the ontological committment to *ficta* that such sentences seem to provide. For this would lead to a dangerous slippery slope as regards similar sentences that patently do not involve that committment. Moreover, there is a reason why the latter sentences do not involve that committment; namely, that they are merely fictionally true. As a result, *contra* (b) the former sentences must be merely fictionally true as well. More formally, this is Kroon's argument (KA):

(i) If we accept that sentences like Van Inwagen's (1)–(2) ontologically commit us to fictional entities, we must accept that other even more problematic existentially quantified sentences ontologically commit us to even more problematic entities;
(ii) But we do not accept the latter committment;
(iii) Hence, we must not accept the former committment as well.
(iv) The latter sentences are just fictionally true;
(v) Hence, so are the former sentences.

To begin with, in order to criticise the first step of KA, a creationist may reject (ii): though bizarre, we welcome the problematic entities. But this move, which Meinongians of any sort may well please, is unavailable to Creationism. A creationist does not want to admit entities of any sort, but just entities that can be legitimately traced back to a process of creation. So, a creationist must

[8] Cf. Van Inwagen (1977). See also his (2000, 2003).

grant (ii). Indeed, she may agree with Kroon that there are existentially quantified sentences that induce no ontological committment. As a matter of fact, they are merely fictionally true. Consider Kroon's own examples:

(3) Some imaginary companions are more closely modelled after playmates of the child's own age, size and gender.
(4) There are all kinds of things that are posited simply as a result of people misapprehending what others say.

Granted, a creationist may wonder whether by (3), which allegedly involves imaginary companions, Kroon provides a good example of an existentially quantified yet ontologically noncommittal sentence. For she may take imaginary companions on a par with fictional entities, hence as abstract entities of some similar sort, thereby considering (3) as a genuinely true sentence that commit one to such entities.[9] But the example Kroon provides in (4), which allegedly involves entities that are even more metaphysically extravagant, is for the creationist noncontroversial in this respect. A creationist hardly wants to be ontologically committed to such pseudoposits: misapprehension hardly leads to the creation of any entity. Indeed, a creationist takes (4) as at most fictionally true.

A creationist may rather put (i) into question, however.[10] This is precisely MC's task. For MC, existentially quantified sentences that are seemingly *ficta*-involving induce an ontologically committment to such entities, just as their genuine truth suggests. But this does not mean that *all* similar sentences induce that committment; at least sentences like (4) actually fail to be genuinely true. Unlike the latter sentences, the former sentences are backed by a reflexive stance of the kind sketched in the previous section that leads to the generation of certain entities; namely, *ficta*.

So, says MC, sentences (1)–(2) are ontologically committal. For there are at least reflexive stances of the above kind on the *per se* ontologically noncommittal make-believe practices lying behind them. Such stances indeed lead to the *ficta*-generation that allow for (1)–(2) to be genuinely true existential quantifications. Yet no such stance occurs as to sentences involving other ontologically problematic cases.

In this respect, (4) is not the only case where no such stance occurs. In a dream, one may utter sentences that, within the scope of the dream, are true, even if the singular terms they contain refer to nothing whatsover (they refer to something unactual, typically concrete, yet just in the scope of the dream). In this respect, oneiric telling practices are uncontroversially ontologically noncommittal, just like make-believe practices. Unlike the latter practices, however, with respect to oneiric telling practices no reflexive stance occurs. In order for that stance to occur, it must be public: not only the dreamer, but also other people, must share it. Yet there is certainly no evidence that, as regarding dreams, people share any such stance.

[9] For an attempt at going in this direction, cf. Zvolenszky (2015b: 182).

[10] For a similar move aimed at blocking the inference from Fictional Creationism to Mythical Creationism, the theory according to which in the overall ontological domain there also are mythological entities often not recognised as such even by their creators, cf. Goodman (2014).

Thus, no existentially quantified sentence even arises out of those practices that is both ontologically committal and genuinely true. At most, any such sentence is just fictionally true, hence it is again non- ontologically committal just as the oneiric telling practice lying behind it.

Suppose e.g. that, after having heard a dreamer mumbling about the utterly unactual yet concrete Siggy and Freudy, 'whom' the dreamer refers to just in the scope of her dream by qualifying 'them' in a way that makes 'them' surrogates of certain of her relatives, a psychoanalyst utters, as a sort of theoretical comment:

(5) Some oneiric entities are more closely modelled after relatives of the dreamer's own age, size and gender.

Despite that utterance, (5) ontologically commits us to no alleged oneiric entity generated out of that dream. So, it is not even genuinely true, but at most fictionally true. Indeed, no reflexive stance has occurred as to that oneiric telling practice. As we have just seen, for MC the reflexive stance having to do with the relevant make-believe practice must be public in order for it to display a normative force that allows the right item – a given *fictum* – to arise out of it. In the case in question, however, there is no reflexive stance, for people *actually* share no such stance.[11] Hence, says MC, there are no oneiric entities.

Note that this criticism of KA's premise (i) on MC's part focuses on the particular *ontological* reason why for MC sentences (1)–(2) are genuinely true: they are such because they are ontologically committal, not the other way around. Thus, unlike premise (a) of VA, the original realist argument by van Inwagen, this criticism of KA's i) does not rely on the genuine truth of sentences (1)–(2) in order to get an ontologically committal conclusion.[12] By the same kind of reasoning, sentences like (4) and (5) are not genuinely true – at most, they are just fictionally true – for they are not ontologically committal.

This way of putting things, moreover, leads MC to obtain a further result; namely, that the second step of KA leading from (iv) to (v) does not even go through. Even if problematic existentially quantified sentences like (4) and (5) are just fictionally true, sentences (1)–(2) are not. Granted, such sentences *may* be given a non- ontologically committal reading in which they are fictionally true. Yet even if there were such a reading this would not be their only reading, as the famous Pirandellean sentence:

(6) There are fictional characters in search of an author

shows. (6) is certainly fictionally true, when uttered within the make-believe game Pirandello inaugurated in writing *Six Characters in Search of an Author*. In that

[11] Certainly, on this respect things might change. We would then have an oneiric entity. If we had something like *oneiric works*, just as we have fictional works, this would be a sign that a public reflexive stance in favour of oneiric entities has occurred.

[12] As a result, even if one managed to independently show (as I incidentally believe it is hard to do: Voltolini 2018) that VA's (a) is false (as many people nowadays hold: Azzouni (2010), Crane (2013), Everett (2013), Sainsbury and Tye (2012)), this would be of no help for the anticreationist.

game, one noncommittally makes believe that there are such characters: there are such characters just in the scope of the game (the game is tricky, for in it one makes believe that there are fictional entities, not concrete individuals as is typically the case). Yet this is not the only way for (6) to be true. For it may be uttered outside of that game as a bit of literary criticism, when it is genuinely true, since it is backed by a reflexive stance that leads to the generation of certain (actually, six) *ficta*: The Father, the Mother, the Stepdaughter, etc.

In (2015), however, Kroon refined his argument so as to skip the previous MC's reply. For, he says, one may apply his argument not to cases that allegedly involve entities of a different categorical kind from *ficta*, such as imaginary companions, oneiric entities, or other even more metaphysically extravagant entities, but to cases that allegedly involve entities of the same metaphysical kind – namely, *ficta* again – which however are ontologically unpalatable. For unlike ordinary *ficta*, these new *ficta* are utterly indiscernible insofar as they are ontologically *indeterminate*. One cannot count how many such entities there are;[13] they are subjects of plural distributive yet non- individually applicable predication.[14] Hence, the relevant existentially quantified sentences are noncommittal; thus, the same must hold as to the Van Inwagen sentences. As a matter of fact, the former sentences are not genuinely true; at most, they are fictionally true. So are the latter ones. More formally, Kroon's refined argument (KRA) runs as follows:

(i_*) If we accept that sentences like the above Van Inwagen's (1)–(2) commit us to fictional entities, we must accept that other even more ontologically problematic existentially quantified sentences commit us to entities of the *same kind*, yet indeterminate;
(ii_*) But we do not accept the latter committment;
(iii_*) Hence, we must not to accept the former committment either.
(iv_*) The latter sentences are just fictionally true;
(v_*) Hence, so are the former sentences.

Consider indeed a sentence like:

(7) There are numerous fictional individuals, including the many dwarves who fought valiantly in the War of the Dwarves and Orcs of Tolkien's *The Lord*

[13]Notoriously, the easiest way for a realist to solve the problem of indeterminate *ficta*, namely, appealing to *groups* as fictional characters such that according to the relevant fiction it is indeterminate how many individuals they comprise (cf. Parsons 1980:191), does not work. For suppose that the following sentence (7) continued by saying "... yet just some of such dwarves appear in many other battles *LOTR* focuses on". It is unclear how appealing to groups would enable a realist to deal with this continuation. For it is unclear whether it would involve the same subgroup or different ones. Analogous problems may be raised by a creationist appeal to similar plural entities like pluralities or sets.

[14]This specification helps Kroon in ruling out a possible creationist reply (that may however be affected by the problem raised in fn. 13) that the relevant existentially quantified sentences are ontologically committal, since they concern either pluralities or sets to which properties are predicated collectively.

of the Rings (LOTR), who are not presented with enough detail to render them distinctive but despite this are more significant to the plot of *LOTR* than many of that work's more visible characters.

In (7), one seems to quantify over indeterminate *ficta* by plurally and distributively, yet nonindividually, predicating properties of them (e.g., *being more significant to the LOTR plot than many of the work's more visible characters*). Yet no creationist would like to take (7) as inducing an ontological committment to utterly indiscernible fictional entities. Indeed, if it is true, it is not genuinely so. Yet if (1)–(2) induce an ontological committment, so does (7). Hence, there is a good reason not to take also (1)–(2) as inducing that committment. Indeed, if (7) is merely fictionally true, so are (1) and (2).

Yet MC may rejoin that once again, neither (i_*) nor (v_*) of KRA are true. True enough, (7) induces no ontological committment; indeed, it is just fictionally true. Yet this depends again on the fact that no reflexive stance occurs that may justify (7)'s being so inducing. A reflexive stance amounts to taking a make-believe practice as involving a certain property set that, in the make-believe, is identical with a certain (nonactual yet typically concrete) individual. Yet in the indeterminacy case there is no such stance, for no one can take that practice in this way. For within that practice there clearly is *no single* (nonactual yet typically concrete) individual that set is identical with. To come back to Kroon's example, in the relevant make-believe practice that is connected to John Tolkien's writing *LOTR*, there are just indeterminately many concrete individuals that are dwarves and fought valiantly in the War of Dwarves and Orcs. So, within that practice there is *no single* concrete individual the set constituted (possibly *inter alia*) by the properties of *being a dwarf* and *having fought valiantly in the War of Dwarves and Orcs* is identical with. Hence, no reflexive stance can affect the practice itself. Granted, the make-believe practice involves concrete yet indeterminate individuals that are dwarves and fought valiantly in the relevant battle. This is not surprising. Just as that practice may describe a world that is neither coherent nor complete, in its being ontologically noncommittal it may also describe a world whose (concrete) individuals are indeterminate. By contrast, in its being ontologically committal the reflexive stance leading to a *ficta* generation must generate a determinate number of *ficta*, however abstract and mind-dependent they are. Thus, the fact that (7) induces no ontological committment does not prevent (1)–(2) from being so inducing. As a result, the fact that (7) is just fictionally true does not mean that (1)–(2) are also such.[15]

[15] In this case, which for some involves a mere *semantic* indeterminacy (the description of the make-believe world is conceptually indeterminate) the make-believe practice undetermines the creation of different indeterminate *ficta* (one and the same set of properties may not be make-believedly identical with just one individual). In other perhaps more radical cases, which involve a *metaphysical* indeterminacy (the make-believe world is indeterminate as to whether one individual is the same as another individual) the practice has not even a start, for in the make-believe world there is *no* (typically concrete) individual the set may be makebelievedly identical with. Cf. a

Granted, this way of telling the case of (7) apart from that of (1)–(2) amounts to an application of the "Divide-and-Rule Strategy" (Kroon 2015: 164–6), which definitely involves a non *ficta*-involving semantic treatment for (7), unlike (1)–(2). Yet *pace* Kroon[16] this application does not rely on semantic resources that can also be mobilised as far as (1)–(2) are concerned, thus leading to an argument of the (KRA)- style. Rather, it relies on an ontological consideration that shows why, unlike (1)–(2), (7) cannot be ontologically committal, thus being at most merely fictionally true.

One may reply that this way of appealing to a reflexive stance is rather ad hoc. The make-believe practice is such that in it, it is indeterminate how many (concrete) individuals there are iff there are indeterminately many *ficta* that are appropriately qualified by properties mobilised in the practice. Thus, it seems that a creationist must swallow the unpalatable ontology of indeterminate *ficta* Kroon rightly recommends not to endorse.[17]

Yet this reply relies on the plausibility of the bridge principle it appeals to: the make-believe practice is such that there are indeterminately many relevant concrete individuals iff there are indeterminately many relevant *ficta*.[18] But why should one endorse this principle? If the world of a make-believe practice is indeterminate[19] as regards the cardinality of the typically concrete individuals it contains, there is no reason as to why this indeterminacy must be mirrored in the cardinality of *ficta* that arise out of that practice. Granted, if there were a one-one correspondence between the pseudodomain of individuals within a make-believe world and the domain of *ficta* that arise out of the relevant make-believe practice, as the 'something from nothing' conditional Radical Creationism appeals to ("If one mobilises the name 'NN' within a make-believe storytelling practice, then 'NN' refers to a fictional entity") suggests, that bridge principle would be justified. Yet for MC that conditional is not true. So, the bridge principle is not justified. In this respect, appealing to the lack of a reflexive stance in order to dispense with the problem of indeterminate *ficta*, rather than being *ad hoc*, provides a reason as to why bridge principles such as the above one must not be endorsed.

Let me end this Section with a general remark. Clearly enough, Kroon's ontological antirealist arguments are addressed against a semantical argument for *ficta*, Van Inwagen's VA, that a realist about such entities may defend even if she is not a creationist, by sticking to thesis (a) above without also sticking to thesis (b). Thus, a creationist may generally retort as follows. Suppose that the previous ontological considerations against Kroon's arguments were ungrounded. Nevertheless, if there

version of Robert Stevenson's tale according to which it were indeterminate whether Dr. Jekyll is identical with Mr. Hyde.

[16] Cf. Kroon (2015: 166).

[17] For similar cardinality problems affecting *ficta*, cf. Everett (2013), Nolan and Sandgren (2014).

[18] Cf. Everett (2013: 226). To my mind (Voltolini 2015a), the problems other similar bridge principles Everett appeals to arise for *ficta* can be dealt with in a similar way.

[19] Either in a semantic or in a metaphysical sense. Cf. fn.15.

were an independent way to support VA, e.g. by supplementing it with a genuine ontological argument in favour of *ficta*, that would just amount to a defense of realism about *ficta*, not of Creationism in particular. For, as we saw before, Creationism is distinctively qualified by thesis (b).[20] Granted, Kroon is skeptical as to whether one may find a genuinely ontological argument in favour of *ficta*. For Kroon, if one holds that a fictional work is individuated by discernible fictional entities,[21] one must implausibly hold as well that such a work is individuated by utterly indiscernible, namely, indeterminate, *ficta*.[22] But if, as argued above, there are no indeterminate *ficta*, a fictional *work* may hardly be individuated by them. While in a make-believe practice one may tell that there are indeterminately many (unactual yet typically concrete) individuals, a fictional work is individuated just by the only *ficta* that there are, namely, *determinate* fictional entities. As Kroon himself admits,[23] as far as the determinacy/indeterminacy issue is concerned, one may simply apply the aforementioned Divide-and-Rule Strategy.

Thus, being able to dismantle critiques to her thesis (a) is not enough for a creationist. For an anticreationist may instead appeal to the *metaphysical* criticisms that are addressed against thesis (b) of Creationism, which, as we saw before, is Creationism's distinctive thesis. It is now time, therefore, to move on to assess those criticisms.

14.4 The Metaphysical Criticisms

Two main criticisms have been addressed against a creationist metaphysics concerning fictional entities. According to Deutsch (1991), on the one hand, the whole idea of a *creation* of a *fictum* is misplaced, at least if we appeal to just one notion of a creation. For according to that notion, creation relies on a causal process – creating in this sense is bringing something into existence – yet no causal process can be involved in the generation of an abstract entity, as creationists suppose *ficta* to be. For *abstracta* are devoid of causal powers. Indeed for Deutsch, saying that *ficta* are created rather means that they are *stipulated* to be what they are like; hence, the very notion of a creation must be given an anticreationist conceptual twist. According to Brock (2010), on the other hand, even if one allowed for *ficta* to have causal powers since their being *abstracta* does not prevent them from *coming* into existence, there is no cogent reason as regards which moment must be taken as the *real* creative moment, the moment in which, in virtue of the relevant imaginative activity, a certain *fictum* comes into existence. For since Creationism does not explain *how*

[20] For examples of such those arguments cf. e.g. Thomasson (1999), Voltolini (2006).

[21] For this idea cf. Voltolini 2006.

[22] Cf. Kroon (2015: 170fn.29).

[23] *Pace* Everett (2005, 2013) and Howell (2011). Cf. Kroon (2015: 165–6).

it comes into existence, it cannot even say *when* it comes into existence. To my mind, however, MC can reject both criticisms.

Let me start with Deutsch's criticism. This criticism is utterly correct if the creation relation is conceived as a *causal* relation. For no causal relation can properly concern an entity that, *qua* abstract, is devoid of causal powers. Granted, one may plausibly say that *ficta* have a causal impact on one's mind, by affecting one's emotions.[24] Yet this hardly is the sort of causality that allows *ficta*, as well as *abstracta* more in general, to have causal powers. For, as Frege originally envisaged,[25] in order for something to properly have causal powers, it must not only prompt but also undergo effects. If an entity has just one of the above two features, as it happens with *ficta* in their merely being able to determine emotions, it properly has no causal powers. Certainly, one may say that by being brought into existence, *ficta* are causally sensitive.[26] Yet it is controversial whether this is enough to ascribe *ficta* the power of undergoing effects, since no further modification can be induced on them; particularly if, as we will immediately see, there is an alternative way of understanding their generation.

Indeed, just as RC, MC may come both in a *weak* and in a *strong* version. While in the weak version the relation between the reflexive stance and the fictional entity that arises out of it is *causal*, in the strong version that relation is *constitutive*. In order to dispense with Deutsch's criticism, a creationist must therefore endorse the strong version of MC, as I have actually done all along.

In point of fact, the 'creation' of a fictional entity conforms with the generation of an institutional entity via the appropriate constitutive rules (Voltolini 2015b). What brings an institutional entity into existence is the application of one of Searle's (1969, 1995) *constitutive* rules; namely, rules that constitute the phenomena they rule. Abiding by such rules is what prompts institutional states of affairs to subsist. In Searle's terms, this normative process makes a certain entity X *count* as another entity Y in the appropriate context C. I italicise "count" in order to stress that the relation between X and Y is not an identity relation: by making an institutional state of affairs come into existence, context C shapes a previously existing entity X into the new entity Y. For example, producing certain utterances is the right way for an individual's being baptised with a certain name. These utterances count, in the relevant context involving the named person herself, certain utterers endowed with the proper authority, and a given community of reference, as that person's being baptised with that very name. Likewise, as regards *ficta* constitutive rules are at stake as well, as Thomasson (2003a, b) originally maintained: abiding by such rules prompts a certain *fictum* into existence. Yet such rules do not involve the relevant make-believe practice as such, as Thomasson thinks, but rather the reflexive stance that allows one to take that practice as involving a certain property set. Indeed, *pace* Thomasson, the conceptual truth that expresses the relevant constitutive rule is not

[24] As Zalta (1988: 4, 127–8) suggested.

[25] Cf. Frege (1997: 370–3).

[26] Cf. Brock et al. (2013).

that if one mobilises the name 'NN' within a make-believe storytelling practice, then 'NN' refers to a fictional entity. As we saw before, antirealists can deny the above conditional and yet share our notion of a *fictum*. Rather, the conceptual truth in question is that if one takes a certain make-believe practice to involve a certain property set as make-believedly identical with a certain (nonactual yet typically concrete) individual, then a certain *fictum* comes into existence (*qua* a certain abstract mind-dependent entity).[27] Insofar as they share our notion of a *fictum*, antirealists cannot deny that *if* there is such a stance, there also is a certain *fictum*.[28] At most, they can deny – wrongly, in my opinion – that there is such a stance – as they would instead correctly do e.g. with respect to the oneiric entities I talked about in the previous section. Why so? For the reflexive stance is the contextual factor that allows the relevant make-believe practice to count as a *fictum*. In Searle's terms, in the context in which a make-believe practice is taken by the people having the proper sensibility (fiction makers, fiction readers) to involve a property set as make-believedly identical with a certain (nonactual, typically concrete) individual, a *fictum* is *eo ipso* generated, as an abstract mind-dependent entity. In endorsing that stance, such people are endowed with the right generative authority insofar as they are socially recognised to have a proper understanding of fiction. The make-believe practice is not the *fictum*, but in that context it counts as that *fictum*, which is the new institutional entity to be properly generated by abiding the relevant constitutive rules. Indeed, taking that practice in the above way makes that *fictum* the *right* entity to be generated. It would be wrong for any other *fictum* to be an entity ontologically correlated with that practice so taken. For it is clearly the practice itself and the set of properties that it mobilises that determine which *fictum* must be generated.[29] Compare this case again with the situation of baptism. Uttering a name in certain conditions counts as the baptised person's acquiring the name, so that the name is the *right* name for the person. For being baptised consists in that person's thereby *having* that name, not just being addressed by means of it.[30]

This way of putting things as regards *ficta* generation is particularly evident if one also holds that such a way of taking the practice singles out elements that are metaphysically relevant for the very identity of a *fictum*: a certain property set and

[27]Clearly enough, this conceptual truth must not be transparent to people involved with make-believe practices. But this is not particularly astonishing. As Wittgenstein (2009)[4] stressed, grammatical propositions, i.e., what fix the rule of use of certain expressions hence their meanings viz. the concepts they express, are to be shown to the community members that abide by them.

[28]*Pace* Everett (2013: 131–2), that our notion of a *fictum* involves the fact that a *fictum* does not exist is clearly accountable by creationists as meaning that a *fictum* is not spatiotemporal.

[29]In this respect, my account bears some similarities with Manning's approach, which claims that "we must correlate fictional objects with the specific *features* through which their native works represent them" (2014: 21). Yet Manning frames this claim within traditional creationist accounts of *ficta* generation, which, as he admits, explain our realist intuitions about *ficta* that however occur independently of whether a realist or antirealist theory about *ficta* is correct (2014: 21–2).

[30]For Fara Graff (2011), this amounts to the difference between *being called N* and *being called 'N'*.

the make-believe practice that applies to it. More precisely, a *fictum* is a hybrid entity make by a certain make-believe practice and by the property set it mobilises. In its being generated, a *fictum* comes into existence at a certain time. Yet since it is constituted also by a (timeless) property set, it is an everlasting entity (Voltolini 2006).

One might now wonder whether Strong MC simply rejects, like Deutsch, the idea that the pertinent notion of a creation of *ficta* is the one involving causation rather than the one involving stipulation. Aren't entities that owe their existence to our abiding by constitutive rules *stipulated* entities? In a certain sense, this is correct. For, in virtue of its being public, the constitutive rule whose following leads to the generation of a *fictum* has a conventional, or a stipulative, character. Yet in another sense it is not. For Deutschian stipulations have a different nature and concern metaphysically different entities. Indeed, Deutsch takes the relevant stipulation to concern the what-it-is-like of a certain *fictum*: it is a construction coinciding with the what-it-is-like of an already existent entity in the overall ontological inventory. Moreover, it is always successful, for it is always the case that a certain entity that is already there matches it.[31] However, not only what is stipulative for Strong MC are the constitutive rules, but also, by appealing to such rules, Strong MC parts company with the Platonist aspect of Deutschian creation. Unlike Deutsch, for Strong MC *ficta* are (abstract and) mind-dependent entities, since they come into existence in virtue of our following certain constitutive rules. So, for Strong MC *ficta* can be seen to be stipulated entities *qua* the outcome of certain constitutive rules; they are not entities that are there *before* the stipulation.

Armed with the above reflections, moreover, Strong MC can dispense with Brock's criticism as well, by proving what Brock (2018) considers an *ambitious* response to the problem he raises to creationism. To recall, Brock's first problem, the *'when'*- problem, is that no moment is the real moment for a *fictum* to come into existence: neither the moment in which a storyteller starts telling her story, thereby inaugurating a certain make-believe practice, nor the moment in which she ends telling such a story, by giving a stop to that practice, and not even some intermediate moment inbetween. Moreover, no such moment can be the real moment, for creationists do not manage to address the *'how'*- problem: they have no convincing story as to *how* a *fictum* comes into existence. In this respect, Brock is absolutely right in claiming that appealing to the storyteller creative intentions is unhelpful. For as he says, the creative intention is insufficient in this respect. For first, it implausibly undergenerates *ficta* (the storyteller does not manage to generate a *fictum*, as she wished: take e.g. someone who, by erroneously believing she is having a hallucination rather than a perception, intends to write a tale about something that, unbeknownst to her, is a real individual). Second, it overgenerates *ficta* (notwithstanding the storyteller's wishes, people do not get

[31] Cf. Deutsch (1991: 222–3). Deutsch appeals to the socalled principle of poetic license (ibid.: 211) that reminds of the Meinongian object-abstraction principle Meinong (1983) labeled the principle of the Freedom of Assumption.

different *ficta*: for example, suppose that two sentences each containing the same description, as been intended by their author to characterize distinct individuals, are inadvertently put together by the author herself, so that such descriptions are read as a mere repetition – as happens with a song's refrain – thereby mobilizing just an individual).[32] Nor is such an intention necessary, for a *fictum* may arise even in the absence of that intention (consider an example opposite to the one I proposed a few lines above: this time the storyteller erroneously believes that she is having a perception instead of the hallucination she actually entertains, thereby erroneously believing that she is telling a true story about a real individual, rather than a fictional story about an unreal individual).[33]

Yet first, as regards the 'when'- problem, Strong MC claims once again that it is not the make-believe practice, but the very reflexive stance on that practice that 'wears the trousers' as far as *ficta* generation is concerned. Indeed, for Strong MC *whenever* that stance occurs, a certain *fictum* is generated. Hence, a certain make-believe practice may flow on and different *ficta* may arise out of it, insofar as different reflexive stances occur with respect to it by focusing on different property sets included in that continuing practice.

This typically happens as regards cycles. Consider e.g. the cycle of mock-heroic poems that Chretien de Troyes' *Chanson de Roland* actually inaugurated. Out of Matteo Maria Boiardo's telling the story of *Orlando in Love*, a certain fictional Roland is generated. Yet even if later on Ludovico Ariosto deliberately (and successfully) intended to let the previous make-believe practice continue by his telling the story of *Orlando Enraged*, a different fictional Roland arises out of that continuation. For the respective reflexive stances that applied to that continuing practice mobilised different property sets (e.g., the property of *being mad* figures in the second, but not in the first, set).[34]

Second, entertaining that stance is the right generative *moment* for it also tells us *how* a *fictum* is generated. Indeed, as we saw before, if one endorses Strong MC, the generation of a *fictum* has to do not with *causal* factors, but with *constitutive* rules.

Granted, on behalf of Brock one may remark that, if the constitutive process is involved in the same causal chain mobilising the imaginative activities that lie behind the generation of a *fictum*, that generation is still causally dependent on

[32]Cf. Brock (2010: 355–62). I have deliberately changed Brock's examples in order not to get into the controversy involving the number of immigrant *ficta*, as Brock instead does. For that controversy may be accounted for differently according to one's theory on crossfictional identity. For a rejection of Brock's use of his own examples cf. Friedell (2016).

[33]Cf. again Brock (2010: 355–62). On the last point see also Zvolenszky (2015b: 181–3) and Brock himself (2018); for a moderate view that distinguishes between intentional creation and nonintentional production of a *fictum*. cf. Cray (2017). Again, I have changed Brock's example, for it implausibly appeals to a storyteller antirealist *philosophical* beliefs, which may be inert as regards a *nonphilosophical* practice of storytelling.

[34]This is not to deny that a general *fictum* may encompass certain particular *ficta*. Cf. Voltolini (2006, 2012).

those activities.[35] Yet once that process amounts to a reflexive stance of the above kind, Strong MC may reply that there is no need to *causally* link that stance with those imaginative activities, which merely remain a necessary condition for that generation.

Let me take stock. If I am right, by endorsing MC in its strong, constitutive, version, a creationist may be able to dispense not only with the ontological criticisms against Creationism, but also with the main criticisms that have been raised against the creationist metaphysics. As a final remark, let me stress again that, certainly, if the relation between a reflexive stance and a *fictum* is constitutive, as Strong MC claims, the very word 'creationism' inappropriately denotes this philosophical position, provided that 'creation' merely signifies a causal process.[36] Rather, as Deutsch stressed, a different notion of creation must be involved that takes into account the stipulative aspect that *ficta* generation mobilises. Indeed, the fact that the word 'artifactualism' is often used instead of 'creationism' to label this metaphysical position on *ficta* is a symptom of this sort of linguistic inappropriateness. But even 'artifactualism' is somehow unsatisfying. *Ficta* do not come into existence in virtue of some manufactural process assembling some already existent items in a new way (typically in accordance with certain architectural intentions), as it happens with concrete artifacts. Certainly, one may say that in the generation of a *fictum* an author assembles properties in a certain way. Yet it is unclear not only whether properties are the right kind of material to be assembled – they are *abstracta* as well – but also whether there is in such a case something like an architectural intention (Voltolini 2006). Indeed, speaking of *ficta* as artefacts does not capture the constitutive rather than the merely stipulative character that a *fictum's* generation must possess. For, unlike Deutsch, a *fictum* is not a pre-existing entity that in its features matches an inventive authorial characterization. Rather, it is the non-preexisting outcome of the application of a certain constitutive rule. Thus, if Strong MC is the right way to support this general position, we are still after a new proper word for it.

Acknowledgments This paper has been presented at the conferences *The Roots of Fiction*, University of Macau, December 5–6 2015, Macau; *BSA Conference on Fictional Characters*, University of Southampton, December 15–16, 2015 Southampton; *Political Communities. Normativity and the Metaphysics of politics* July 2–3, 2019, University of Udine, Udine. I thank all the participants for their useful remarks. I also thank Catherine Abell, Fred Kroon and Elisa Paganini for their insightful comments to previous version of the paper.

References

Azzouni, J. (2010). *Talking about nothing: Numbers, hallucinations, and fictions*. Oxford: Oxford University Press.

[35] Cf. Brock et al. (2013: 72).

[36] As Brock (2010: 343) holds *contra* Deutsch (1991).

Berto, F. (2013). *Existence as a real property*. Dordrecht: Springer.
Braun, D. (2005). Empty names, fictional names, mythical names. *Noûs, 39*, 596–631.
Brock, S. (2010). The creationist fiction: The case against creationism about fictional characters. *Philosophical Review, 119*, 337–364.
Brock, S. (2018). A recalcitrant problem for abstract creationism. *Journal of Aesthetics and Art Criticism, 76*, 93–98.
Brock, S., Maslen, C., & Ngai, J. (2013). A puzzle about fictional characters. In C. Barbero et al. (Eds.), *From fictionalism to realism: Fictional and other social entities* (pp. 63–86). Cambridge: Cambridge Scholars.
Castañeda, H.-N. (1989). *Thinking, language, and experience*. Minneapolis: University of Minnesota Press.
Crane, T. (2013). *The objects of thought*. Oxford: Oxford University Press.
Cray, W. (2017). Abstract generationism: A response to Friedell. *Journal of Aesthetics and Art Criticism, 75*, 289–292.
Deutsch, H. (1991). The creation problem. *Topoi, 10*, 209–225.
Evans, G. (1982). *The varieties of reference*. Oxford: Clarendon Press.
Everett, A. (2005). Against fictional realism. *Journal of Philosophy, 102*, 624–649.
Everett, A. (2013). *The nonexistent*. Oxford: Oxford University Press.
Fara Graff, D. (2011). You can call me 'stupid', ... just don't call me stupid. *Analysis, 71*, 492–501.
Frege, G. (1997). The thought. In M. Beaney (Ed.), *The Frege reader* (pp. 325–345). Oxford: Blackwell.
Friedell, D. (2016). Abstract creationism and authorial intention. *Journal of Aesthetics and Art Criticism, 74*, 129–137.
Goodman, J. (2004). A defense of creationism in fiction. *Grazer Philosophische Studien, 67*, 131–155.
Goodman, J. (2014). Creatures of fiction, objects of myth. *Analysis, 74*, 35–40.
Howell, R. (2011). Fictional realism and its discontents. In F. Lihoreau (Ed.), *Truth in fiction* (pp. 153–202). Frankfurt: Ontos.
Ingarden, R. (1973). *The literary work of art*. Evanston: Northwestern University Press.
Kripke, S. (1980). *Naming and necessity*. Oxford: Blackwell.
Kripke, S. (2011). Vacuous names and fictional entities. In *Philosophical troubles: Collected papers* (Vol. 1, pp. 52–74). Oxford: Oxford University Press.
Kripke, S. (2013). *Reference and existence*. Oxford: Oxford University Press.
Kroon, F. (1992). Was Meinong only pretending? *Philosophy and Phenomenological Research, 52*, 499–527.
Kroon, F. (2011). The fiction of creationism. In F. Lihoreau (Ed.), *Truth in fiction* (pp. 203–222). Ontos: Frankfurt.
Kroon, F. (2013). The social character of fictional entities. In C. Barbero et al. (Eds.), *From fictionalism to realism: Fictional and other social entities* (pp. 87–110). Cambridge: Cambridge Scholars.
Kroon, F. (2015). Creationism and the problem of indiscernible fictional objects. In S. Brock & A. Everett (Eds.), *Fictional objects* (pp. 139–171). Oxford: Oxford University Press.
Lewis, D. (1983). Truth in fiction. In *Philosophical papers I.* (pp. 261–280). Oxford: Oxford University Press.
Manning, L. (2014). Real representation of fictional objects. *The Journal of Aesthetics and Art Criticism, 72*, 13–24.
Meinong, A. (1983). *On assumptions*. Berkeley: University of California Press.
Nolan, D., & Sandgren, A. (2014). Creationism and cardinality. *Analysis, 74*, 615–622.
Parsons, T. (1980). *Nonexistent objects*. New Haven: Yale University Press.
Predelli, S. (2002). 'Holmes' and Holmes. A millian analysis of names from fiction. *Dialectica, 56*, 261–279.
Priest, G. (2016)[2]. *Towards non-being: The logic and metaphysics of intentionality*. Oxford: Oxford University Press.
Sainsbury, M. R. (2010). *Fiction and fictionalism*. London: Routledge.

Sainsbury, M. R., & Tye, M. (2012). *Seven puzzles of thought and how to solve them*. Oxford: Oxford University Press.
Salmon, N. (1987). Existence. *Philosophical Perspectives, 1*, 49–108.
Salmon, N. (1998). Nonexistence. *Noûs, 32*, 277–319.
Schiffer, S. (1996). Language-created language-independent entities. *Philosophical Topics, 24*, 149–166.
Schiffer, S. (2003). *The things we mean*. Oxford: Clarendon Press.
Searle, J. R. (1969). *Speech acts*. Cambridge: Cambridge University Press.
Searle, J. R. (1979). The logical status of fictional discourse. In P. A. French et al. (Eds.), *Contemporary perspectives in the philosophy of language* (pp. 233–243). Minneapolis: University of Minneapolis Press.
Searle, J. R. (1995). *The construction of social reality*. New York: Free Press.
Thomasson, A. L. (1999). *Fiction and metaphysics*. Cambridge: Cambridge University Press.
Thomasson, A. L. (2003a). Fictional characters and literary practices. *British Journal of Aesthetics, 43*, 138–157.
Thomasson, A. L. (2003b). Speaking of fictional characters. *Dialectica, 57*, 205–223.
Van Inwagen, P. (1977). Creatures of fiction. *American Philosophical Quarterly, 14*, 299–308.
Van Inwagen, P. (2000). Quantification and fictional discourse. In A. Everett & T. Hofweber (Eds.), *Empty names, fiction and the puzzles of non-existence* (pp. 235–247). Stanford: CSLI Publications.
Van Inwagen, P. (2003). Existence, ontological commitment, and fictional entities. In M. J. Loux & D. W. Zimmerman (Eds.), *Oxford handbook of metaphysics* (pp. 131–157). Oxford: Oxford University Press.
Voltolini, A. (2006). *How Ficta follow fiction*. Dordrecht: Springer.
Voltolini, A. (2009). The seven consequences of creationism. *Metaphysica, 10*, 27–48.
Voltolini, A. (2012). Crossworks 'identity' and intrawork$_*$ identity of a fictional character. *Revue internationale de philosophie, 66*, 561–576.
Voltolini, A. (2015a). Anthony Everett, *The nonexistent*. *Dialectica, 69*, 611–620.
Voltolini, A. (2015b). A suitable metaphysics for fictional entities. In S. Brock & A. Everett (Eds.), *Fictional characters* (pp. 129–146). Oxford: Oxford University Press.
Voltolini, A. (2018). Can one refer to and quantify over nonexistent objects of hallucination? In P. Stalmaszczyk (Ed.), *Objects of inquiry in philosophy of language and linguistics* (pp. 97–115). Berlin: Peter Lang.
Walton, K. L. (1990). *Mimesis as make-believe*. Cambridge, MA: Harvard University Press.
Wittgenstein, L. (2009)[4]. *Philosophical investigations*. Oxford: Blackwell.
Yagisawa, T. (2001). Against creationism in fiction. *Philosophical Perspectives, 15*, 153–172.
Zalta, E. N. (1983). *Abstract objects*. Dordrecht: Reidel.
Zalta, E. N. (1988). *Intensional logic and the metaphysics of intentionality*. Cambridge, MA: MIT Press.
Zvolenszky, Z. (2015a). An argument for authorial creation. *Organon F, 22*, 461–487.
Zvolenszky, Z. (2015b). Inadvertent creation and fictional characters. *Organon F, 22*(Suppl), 169–184.

Part V
Fictionalism or Realism in Moral Philosophy and Philosophy of Arts

Chapter 15
Moral Folkism and the Deflation of (Lots of) Normative and Metaethics

Mark Balaguer

Abstract In this paper, I do two things. First, I argue for a metaethical view that I call moral folkism. The two main subtheses of moral folkism are as follows: (A) if there are any wrong-like properties, then there's a vast plurality of them; e.g., there's a property of Kant-wrongness, and Mill-wrongness, and Moore-wrongness, and so on; and (B) which of these properties count as genuine kinds of wrongness (i.e., real moral wrongness)—if any of them do—is determined by facts about us, in particular, our usage, intentions, and practices concerning moral words. Second, I discuss the consequences of moral folkism. In particular, I argue that (i) moral folkism leads us to the deflationary conclusion that many of the normative and metaethical questions that philosophers discuss are settled by empirical facts about what ordinary folk happen to mean by their words—and so they're not settled by mind-independent facts about reality. In addition, I also argue that (ii) moral folkism does not imply that applied ethical questions are settled by facts about folk meaning, and (iii) moral folkism does not imply moral anti-realism (i.e., moral folkism is perfectly compatible with a robust sort of moral realism).

Keywords Moral pluralism · Moral realism · Deflationism · Platonism · Ordinary language

Thanks are due to Rebecca Chan, Justin Clarke-Doane, Matti Eklund, and Michaela McSweeney for providing feedback on earlier drafts of this paper.

M. Balaguer (✉)
Department of Philosophy, California State University, Los Angeles, CA, USA
e-mail: mbalagu@exchange.calstatela.edu

J. L. Falguera, C. Martínez-Vidal (eds.), *Abstract Objects*, Synthese Library 422,
https://doi.org/10.1007/978-3-030-38242-1_15

15.1 Introduction

In this paper, I'll argue for a view of moral properties—a view I'll call *moral folkism*—that leads to a deflationary view of many normative and metaethical questions (but not applied ethical questions).[1] I'll provide a complete formulation of moral folkism in Sect. 15.5; for now we just need the first two parts of the view:

(A) If there are any wrong-like properties, then there's a vast plurality of them; e.g., there's a property of *Kant-wrongness* (i.e., *violating the categorical imperative*, or some such thing), and *Mill-wrongness* (i.e., *not maximizing happiness*, or some such thing), and *Moore-wrongness*, and so on; we can call these properties $wrongness_1$, $wrongness_2$, $wrongness_3$, etc.

(B) Which of these properties counts as *wrongness* (i.e., *real moral wrongness*)—if any of them do—is determined by facts about *us*, in particular, our usage, intentions, and practices concerning moral words like 'wrong' (or what we *have in mind* when we use these words). E.g., if it's built into our usage, intentions, and practices that 'wrong' expresses $wrongness_1$, then that makes it the case that 'wrong' *does* express $wrongness_1$ (and, hence, that $wrongness_1$ counts as *wrongness*); and if it's built into our usage, intentions, and practices that 'wrong' expresses $wrongness_2$, then that makes it the case that 'wrong' does express $wrongness_2$ (and that $wrongness_2$ counts as *wrongness*); and so on.

I'll argue for (A) and (B) in Sects. 15.2 and 15.3—or, more precisely, I'll argue for (B) and explain how I would argue for (A) if I had more space. To simplify things, I'll assume in Sects. 15.2 and 15.3 that the following two claims are true:

(i) We use 'wrong' in a property-ascribing way, so that it's *supposed to* express a property; and (ii) our usage and intentions concerning 'wrong' zero in on a *unique* property, as opposed to *no* property or *many* properties. (I'm aware that (ii) is implausible; but assuming it will simplify things, and no harm will come of this.)

In Sect. 15.4, I'll drop these two assumptions, and I'll point out that arguments similar to the one I use to motivate thesis (B) can be used to show that the question of whether (i) and (ii) are true is *also* determined facts about us. In Sect. 15.5, I'll argue that if the arguments of Sects. 15.1, 15.2, 15.3, and 15.4 are correct, then we're led to the deflationary result that many of the normative and metaethical questions that philosophers discuss are settled by empirical facts about what ordinary folk happen to mean by their words—and so they're *not* settled by mind-independent facts about reality. In Sect. 15.6, I'll argue that my view *doesn't* imply that applied ethical questions are settled by facts about folk meaning. And in Sects. 15.7 and 15.8, I'll argue that my view doesn't imply moral anti-realism.

[1]Moral folkism is similar in certain ways to Frank Jackson's (1998) view, but it's also different in important ways.

15.2 The Vast Plurality of Wrong-Like Properties

In this section, I'll indicate how I would argue for thesis (A) if I had more space. Let an *abstract object* be a non-physical, non-mental, non-spatiotemporal object. Let *platonism* be the view that there *are* abstract objects. And let *plenitudinous platonism* (or for short, *PP*) be the view that there's a plenitudinous realm of abstract objects; i.e., it's the view that there exist abstract objects of all possible kinds, or that all the abstract objects that could exist actually do exist. Given this, the argument for (A) proceeds as follows:

> (A1) If there are any wrong-like properties, then they're abstract objects (and, hence, platonism is true); but (A2) if platonism is true, then PP is true; and (A3) if PP is true, then there's a vast plurality of wrong-like properties. Therefore, (A) if there are any wrong-like properties, then there's a vast plurality of them.

The argument for (A1) is based on the empirical claim that our talk of wrong-like properties (i.e., things like *wrongness*, *Kant-wrongness*, *Moore-wrongness*, etc.) is best interpreted as being about (or at least purporting to be about) abstract objects. I think there are extremely strong arguments for this claim, but I don't have the space to rehearse them here; I'm just going to take it as a working assumption that (A1) is true.

(Let me make two disclaimers. First, I don't think ordinary moral claims like 'Eating meat is wrong' are about abstract objects; I'm claiming only that sentences about *wrong-like properties*—e.g., '*Wrongness* is a non-natural property'—are about abstract objects. Second, I'm not claiming that abstract objects actually *exist*; I'm just claiming that sentences about wrong-like properties *purport* to be about abstract objects—so that if wrong-like properties exist at all, then they're abstract objects. But despite this, I think there are ways to avoid endorsing platonism. I don't think this is very important here, though, because the anti-platonist views I have in mind lead to all the same conclusions about morality that platonism leads to. But I can't get into this here.)

The argument for (A2) is based on the claim that PP is the only tenable version of platonism. There are multiple arguments for this. One quick argument is based on the claim that non-plenitudinous versions of platonism involve unacceptable kinds of metaphysical arbitrariness—e.g., they entail claims like '*Blueness* exists, but *redness* doesn't.' But the best argument for (A2) is based on the fact that PP is the only version of platonism that can be given an acceptable *epistemology*. I've argued elsewhere (1998) that if platonists endorse PP, then they can explain how we humans—naturalistic, spatiotemporal creatures that we are—could acquire knowledge of (acausal, non-spatiotemporal) abstract objects. If we focus on mathematical objects, the explanation proceeds roughly as follows:

> Since PP says that there are abstract objects of all possible kinds, it entails that every purely mathematical theory that *could* be true—i.e., that's internally consistent—accurately describes some collection of actually existing abstract objects. Thus, it follows from PP that in order to acquire knowledge of abstract objects, all we have to do is come up with a consistent purely mathematical theory (and know that it's consistent). This is because,

> again, according to PP, *every* consistent purely mathematical theory accurately describes a collection of abstract objects. But if all we need to do to acquire knowledge of abstract objects is come up with a consistent mathematical theory (and know that it's consistent), then we *can* do this. For (a) we *are* capable of formulating consistent mathematical theories (and knowing that they're consistent), and (b) being able to do this doesn't require us to have any information-gathering contact with the abstract objects that the theories in question are about. (Here's an example of how this works: if anti-platonists ask how we could know that, e.g., every natural number has a successor, the PP-ist response is that while there *are* structures in which some number-like things don't have successors, we just *stipulate* that we're not talking about such structures when we do arithmetic; we stipulate that we're talking about a structure in which every number just *does* have a successor. In short, we stipulate that we're talking about a structure that just *is* characterized by the standard axioms of arithmetic—or by our full conception of the natural numbers.)

This is very quick, and there are obvious objections you might raise. I say more about this in my (1998); I also argue there that *non*-plenitudinous platonism *can't* be given an adequate epistemology. But I can't develop the argument in any more detail here; I'm just going to assume that PP is the only tenable version of platonism and, hence, that (A2) is true—but I'll say more about the PP-ist epistemology in Sect. 15.3.

Finally, (A3) is trivial—because PP just straightforwardly entails that there's a vast plurality of wrong-like properties. According to PP, every property that we can dream up (or that *could* exist) actually *does* exist. So, e.g., there are ordinary properties like *redness*, but there are also weird properties like *being a car or a mouse*, and uninstantiated properties like *being a polka-dotted skyscraper*, and indeed, there are even properties that *couldn't* be instantiated, like *being a round square*. So, obviously, according to PP, all of the wrong-like properties that we can dream up—e.g., *Kant-wrongness* and *Mill-wrongness* and so on—exist.

It's important to note that properties of the kind I'm talking about have internal structure, or *decompositional* structure, and they're individuated in a very fine-grained way; so, e.g., *being a round square* is different from *being a round triangle*.

Now, of course, you might think that *wrongness* is a *primitive* property—i.e., that it doesn't have any decompositional structure (or definition, or whatever). That's fine. According to PP, this is just another wrong-like property—i.e., it's one of the properties that's a candidate for being identical to *wrongness*.

You might wonder whether all of the wrong-like properties are *normative*. PP provides an answer to this question. Consider, e.g., the property *not maximizing happiness*. According to PP, this property exists, and so do the following "normativized" versions of it: (i) *not maximizing happiness and, because of this, being such that it ought not to be done*; and (ii) *not maximizing happiness and, because of this, being such that people have a reason not to do it.* According to PP, there's a property corresponding to every way of normativizing the original (un-normativized) property. And whether the original property counts as "wrong-like" won't matter here. (Likewise, PP entails that there's a plurality of *non-natural* wrong-like properties.)

(Disclaimer: while PP implies the *existence* of these properties, it doesn't imply that they're all *instantiated*. It's obvious that *not maximizing happiness* is instantiated, but it's not obvious that normativized versions of it are.)

15.3 Conceptual Analysis

In this section, I'll assume that we use 'wrong' in a property-ascribing way (and that our usage and intentions zero in on a *unique* wrong-like property), and I'll argue for thesis (B)—i.e., roughly, for the following claim:

> Which of the various wrong-like properties counts as *wrongness* is determined by facts about *us*—about our usage, intentions, and practices concerning words like 'wrong', or about what we have in mind when we use these words.

We can think of thesis (B) as an answer to a certain metaphilosophical question. To see what I've got mind here, consider the following two questions:

> *The what-is-wrongness question*: What is *wrongness*? I.e., assuming that 'wrong' expresses a property, which wrong-like property does it express? Or, since it's trivial that 'wrong' expresses *wrongness*, we can put the question like this: Which of the various wrong-like properties counts as *wrongness*?
>
> *The metaquestion*: What kinds of facts determine the answer to the what-is-wrongness question?; more precisely, if we assume that 'wrong' expresses a property, what kinds of facts determine which property (or properties) it expresses?

Thesis (B) is an answer to the metaquestion: it tells us that the answer to the what-is-wrongness question is determined by facts about *us*—about what we mean by 'wrong'. Let's call this view, or this answer to the metaquestion, *the ordinary-language view*. I want to argue for this view—and, hence, for thesis (B)—by arguing that no other answer to the metaquestion is plausible. Other answers to the metaquestion say that other kinds of facts—aside from facts about folk meaning—are relevant to determining the answer to the what-is-wrongness question. (Presumably, the idea here is that there are moral or metaphysical facts—that are independent of us—that make it the case that one of the wrong-like properties is *privileged* in some way.) In this section, I'll argue that there are no such facts.

You might think I'm conflating two different questions here, namely,

> *The semantic question*: What property is expressed by 'wrong'?; and
>
> *The metaphysical question*: What is the nature (or the decompositional structure) of *wrongness*?

And you might object to my stance by saying something like this:

> The *semantic* question is obviously settled by facts about us; after all, we might have used 'wrong' to express some other property. But this is irrelevant to the *metaphysical* question; that's a question about the nature of a certain abstract object, and it has nothing to do with us.

My response to this objection will emerge in Sect. 15.3.1.

15.3.1 Metaphysical Privilege I—Platonistic Privilege

Picking up on the objection just articulated, you might endorse the following view:

> *The platonistic answer to the metaquestion*: When we acquire the concept *wrongness*, and learn to apply the term 'wrong', we so to speak "grab hold" of a certain property; and we do this without learning the exact nature (or decompositional structure) of that property—i.e., without learning a definition of 'wrong'. If we want to figure out the nature (or decompositional structure) of *wrongness*—whether it's a utilitarian property or a Kantian property or whatever—we need to do some conceptual analysis. And when we do this, we're uncovering facts about a certain abstract object—namely, *wrongness*. And these platonistic facts (about the nature of *wrongness*) are the facts that settle the what-is-wrongness question—or at any rate, the metaphysical (i.e., non-semantic) half of that question. This is analogous to mathematics, where platonistic facts about mathematical objects determine the answers to mathematical questions.

This view is misguided. Given PP, we know that *all* of the various wrong-like properties exist. Moreover, given the PP-ist epistemology described above, we can know what these properties are like without doing any conceptual analysis. E.g., we can know what *Mill-wrongness* is like because when we talk about *Mill-wrongness*, we're just *stipulating* that we're using '*Mill-wrongness*' to denote the property that just *does* have the nature (or structure) that Millians have in mind. And the same goes for *Kant-wrongness* and *Foot-wrongness* and every other wrong-like property that we might be interested in.

So, given PP, we can know what all of the relevant wrong-like properties are like. But this leaves the what-is-wrongness question *open*; after we describe the decompositional structures of the various wrong-like properties we're interested in, there's still the question of which of these properties counts as *wrongness*. So the platonistic answer to the metaquestion is implausible because we can know what all the relevant abstract objects are like—via the PP-ist epistemology outlined above—without having any clue how to answer the what-is-wrongness question. To answer that question, we need to determine which of these abstract objects—whose decompositional structures we already understand perfectly well via the PP-ist epistemology—counts as *wrongness*. And the problem is that the platonistic facts seem entirely *irrelevant* to this question. Indeed, *prima facie*, these considerations seem to lead us right back to the ordinary-language view; they seem to suggest that the answer to the what-is-wrongness question is determined by facts about *us*—about which of the various abstract objects (or decompositional structures) we have in mind when we think and talk about *wrongness*. If, e.g., we have $\mathit{wrongness}_{17}$ in mind when we talk about *wrongness*, then that *makes it the case* that that's the property that we've "grabbed hold of"—and, hence, it's the one that counts as *wrongness*.

You might object as follows: "We can't know what *all* the relevant properties are like in this way because we can't know what *wrongness* is like in this way."

My response: We, in fact, *can* know what all the relevant properties are like in this way; for *wrongness* isn't an *extra* property, on top of all the other wrong-like properties—it's just identical to one of them. The thing we *can't* know in this way is *which* of the wrong-like properties counts as *wrongness*. But this is the question that's settled by facts about us.

Objection: "But after we answer the semantic question—after we figure out which wrong-like property is picked out by our word 'wrong'—we can go on to ask what that property is *like*; and *that* question is a metaphysical question about the nature of an abstract object."

My response: (i) the only way to answer the semantic question (i.e., the only way to specify which wrong-like property is expressed by 'wrong', or which wrong-like property counts as *wrongness*) is to characterize the decompositional structure of the relevant property (e.g., we'd have to say, "'Wrong' expresses the property *violating the categorical imperative*," or some such thing); and (ii) once we've done this, there's no metaphysical question left to answer—we will already have specified the decompositional structure of *wrongness*.

Here's another way to think about this. The *metaphysical* fact behind the right answer to the what-is-wrongness question is a trivial identity fact. Suppose, e.g., that the right answer is that *wrongness* is *Kant-wrongness*. This is just an identity fact. It's interesting to us only because we have two different expressions that denote this property—namely, '*wrongness*' and '*Kant-wrongness*' (or '*violating the categorical imperative*', or whatever). So this is analogous to sentences like 'Hesperus is Phosphorus'; the metaphysical facts behind these sentences are trivial; they're interesting only because (a) we have two different names of the relevant object, and (b) it can be non-obvious that the two names denote the same object. So it's only because of *us*—because we have multiple ways of denoting and expressing the same property—that the sentence '*Wrongness* is *Kant-wrongness*' is interesting (if it's true). And the fact that's non-obvious here—the fact that we need to discover—is a semantic fact.

Note, too, that *all* of the wrong-like properties are identical to themselves. This, of course, is obvious; but it helps bring out the point that the identity fact that we're concerned with isn't *special*; and it isn't interesting *in itself*; it's only interesting because we have multiple ways of picking out the property of *wrongness*.

The overall point, then, is that, vis-à-vis the platonistic facts, there's nothing special about *wrongness*—nothing that makes it stand out from the other wrong-like properties as privileged in a platonistic way. Moreover, the task of uncovering the decompositional structure of this property just collapses into the task of discovering which decompositional structure we have in mind when we talk about *wrongness*. In other words, the so-called metaphysical question collapses into the semantic question.

There's an analogy with mathematics here. We can know what certain mathematical structures are like by just stipulating which kinds of structures we're talking about and appealing to PP to obtain the result that structures of the given kind actually exist. E.g., we can know what standard and non-standard models of arithmetic are like by stipulating that 'standard models of arithmetic' and 'non-standard models of arithmetic' denote structures of certain specific kinds. But this doesn't tell us which of these structures counts as *the natural numbers*. According to the version of PP I favor, this is determined by facts about *us*—about what *we have in mind* when we talk about the natural numbers.

This is related to an important point about mathematical truth. I pointed out above that PP implies that all consistent purely mathematical theories accurately characterize collections of abstract objects. But it doesn't follow that all such theories are *true*, and according to the PP-ist view that I favor, they're *not*. Consider, e.g., the following theory:

NSA: Some natural numbers have infinitely many predecessors.

NSA accurately characterizes certain kinds of abstract objects—namely, non-standard models of arithmetic; but if NSA is put forward as a theory of the natural numbers, then it's *false*, not true. This is because mathematical truth is defined in terms of truth in *intended* structures; NSA isn't true because it's not true of the *intended* objects—i.e., the natural numbers.

So the PP-ist epistemology for mathematics is a bit more complicated than I let on above. There are actually two different kinds of things we can know here. First, we can know what specific abstract objects are like by just stipulating which objects we're talking about and appealing to the plenitudinous nature of the platonic realm (and the consistency of our stipulations) to give us the result that objects of the kinds we're talking about actually exist—and then, if we like, by proving theorems about those objects. And second, we can know which abstract objects count as *the natural numbers* (or *the sets*, or whatever) by getting clear on what's implied by our own intentions, i.e., by what we have in mind when we use expressions like 'natural number' and 'set'. Finally, it's worth noting that the most obvious way to proceed on this second task is to rely on our *intuitions*; this is reliable because our intuitions are generated by our intentions, or by what we have in mind, and so they're windows into what's implied by our intentions. Thus, e.g., the fact that we have an intuition that every natural number has finitely many predecessors is evidence that we have standard (and not non-standard) models in mind when we do arithmetic.[2]

Analogous remarks can be made about conceptual-analysis questions—i.e., questions like 'What is free will?', 'What is knowledge?', and 'What is *wrongness*?' According to PP, we can know what specific wrong-like properties are like by just stipulating which properties we're talking about and appealing to the plenitudinous nature of the platonic realm to get the result that these properties actually exist. And we can know which wrong-like property counts as *wrongness* by figuring out which one we have in mind when we use expressions like 'morally wrong'. And the most obvious way to proceed on this second task is to rely on our *intuitions*—which,

[2]Our intentions might sometimes be imprecise. E.g., our intentions concerning 'set' might not be precise enough to zero in on a unique structure up to isomorphism. If so, there will be some set-theoretic sentences that are true in some intended structures and false in others; on my view (see, e.g., my (2009)), these sentences would be indeterminate—i.e., neither true nor false.

again, is a reliable way of proceeding because our intuitions are generated by what we have in mind when we use expressions like 'wrong'.[3]

I don't mean to suggest that there are no important differences between mathematics and conceptual analysis. One obvious difference is that whereas mathematicians are more centrally concerned with the first task (i.e., discovering the nature of specific abstract objects), conceptual analysts are more centrally concerned with the second task (i.e., determining which abstract objects are picked out by our intentions). My point is just that the epistemologies of the two practices are analogous. In both cases, there are two sorts of things we can know—facts about abstract objects and facts about *us*, about which abstract objects we have in mind when we use certain expressions.

You might object as follows: "On your view, *wrongness* could have been completely different; if we'd just had different thoughts, it could have turned out that, e.g., playing chess was wrong." But this is false; the problem with this objection is that predicates like 'red' and 'wrong' are *rigid*—they express the same properties in all worlds (or all worlds in which those properties exist). So while we could have used 'wrong' to express a different property, it's not true that *wrongness* could have been a different property, or that it could have had a different decompositional structure.

Before moving on, I want to consider one more way in which you might think that platonistic facts are relevant to the what-is-wrongness question. You might think that one of the wrong-like properties is "glowing" somehow in platonic heaven. In other words, you might think that in addition to the kinds of platonistic facts I've been talking about, there are other kinds of platonistic facts that privilege certain abstract objects. E.g., you might think that standard models of arithmetic are "glowing" somehow, or metaphysically privileged, in a way that non-standard models aren't; and you might think that one of the wrong-like properties is "glowing" or privileged in a way that the others aren't.

There are multiple problems with this view. First, it's totally unclear what the "glow" could consist in. (Imagine someone saying that one of the red-like properties, in the spectrum of color properties, is "glowing"; it's entirely unclear what this could mean, and I think it's equally unclear what it could mean to say that one of the wrong-like properties is "glowing".) Second, PP entails that if it's even *possible* for *Mill-wrongness* and *Kant-wrongness* and so on to "glow" in the relevant way, then there are versions of these properties that *do* "glow" in this way; but if this is true, then the facts about which properties are "glowing" won't do the work they're supposed to do. These considerations suggest that what "glow"-platonists are really doing is abandoning PP, and this brings out another problem with the view—it can't be given an acceptable epistemology (in particular, we wouldn't have any way of knowing which abstract objects were "glowing"). Finally, why should we think that the relevant "glow" would be morally relevant? What if

[3] As with the mathematical case, our intentions concerning 'wrong' could be imprecise and, hence, fail to zero in on a unique property. I'll say more about this in Sect. 15.4.

some monstrous kind of *egoist-wrongness* was the one that was "glowing"? Would that mean that that property was *wrongness* and that we should endorse egoism? I don't see why we should think that.[4] Analogy: if it turned out that non-standard models of arithmetic (and not standard models) were "glowing", it wouldn't follow that they were the natural numbers; instead, it would follow that the "glow" was arithmetically irrelevant. And the same seems true in the moral case; if the "glow" didn't line up with our concept of *wrongness*, then it would be morally irrelevant (i.e., it wouldn't be a *moral* glow)—which suggests that what's really doing the work here is our concept of *wrongness*.

15.3.2 Interlude

The remarks of Sect. 15.3.1 suggest that the metaphysical question (i.e., 'What is the nature of *wrongness*?') collapses into the semantic question (i.e., "Which property does the word 'wrong' express?"). This suggests that facts about *us*—about what we have in mind when we talk about *wrongness*—are at least *among the facts* that determine the answer to the what-is-wrongness question. But it doesn't follow that these are the *only* facts that are relevant here, and you might think that other facts—aside from facts about us—are also relevant. However, I will now argue that this is not the case. I'll do this by running through the most obvious facts that one might appeal to here and arguing that they're not relevant to the what-is-wrongness question.

15.3.3 Metaphysical Privilege II—Rigidity and Semantic Externalism

You might think that 'wrong' is a rigidly designating natural-kind term that expresses whatever wrong-like property is actually instantiated in our environment (or whatever wrong-like property causally regulates our usage of 'wrong').[5] Thus, you might think that facts about which wrong-like properties are instantiated in our environment (or which ones causally regulate our usage) are relevant to the what-is-wrongness question.

I've argued elsewhere (2016) that environmental facts of this kind are never relevant to conceptual-analysis questions, but in the present context, this doesn't matter. For there are obviously *lots* of wrong-like properties instantiated in our environment (e.g., there are actions that violate the categorical imperative, and that don't maximize happiness, and so on). Moreover, which of these properties causally

[4] Dasgupta (2017) makes a similar point.

[5] See, e.g., Boyd (1988) and Brink (1989).

regulate our usage depends on facts about *us*, about which of them we focus on and respond to. (It's not as if one of these properties reaches out from our environment and forces us to respond to it in the appropriate way; we can presumably focus on and respond to any wrong-like property that's instantiated in our environment.) So the only thing that environmental facts of this kind could do in the present context is to rule out wrong-like properties that are uninstantiated.

15.3.4 Metaphysical Privilege III—Lewis-Sider-Style Naturalness

One way to respond to what I just said in Sect. 15.3.3 is to claim that the wrong-like property that's the most *natural*, or that does the best job of *carving reality at the joints*, counts as *wrongness*.[6] Elsewhere (2016) I've argued for the general claim that facts about naturalness, or joint-carvingness, are not relevant to conceptual-analysis questions. But in the present context, it doesn't matter whether this general claim is true because in the specific case we're concerned with, it's implausible to suppose that one of the wrong-like properties stands out as more natural than the others.

This, anyhow, is true if we think of naturalness in the way that Lewis (1986) and Sider (2011) do—as having to do with *resemblance*—and if we focus on properties that are plausible candidates for being *wrongness*. Consider, e.g., *Kant-wrongness* and *Mill-wrongness*; it seems obvious that actions that violate the categorical imperative resemble each other to roughly the same degree that actions that don't maximize happiness do; and the same goes for actions that are Foot-wrong and Moore-wrong and so on.[7]

Now, I suppose you might think that there are moral joints in reality that somehow sit on top of the natural facts and make one of the wrong-like properties metaphysically special. But I would respond to this in the same way that I responded to "glow"-platonism: it's unclear what these supernatural joints could consist in; it's unclear how we could know anything about them; and it's unclear why they would be morally relevant.[8]

[6] Dunaway and McPherson (2016) endorse a view like this, and McDaniel (2017) endorses a related view. Williams (2018) argues against views of this kind.

[7] Remarks in this vicinity have been made by Schroeter and Schroeter (2013) and Eklund (2017).

[8] Dunaway and McPherson (2016) claim that the most natural (or "elite") moral properties are the ones that feature in our best moral theories; I'll respond to the appeal to theoretical role in Sect. 15.3.7.

15.3.5 Metaphysical Privilege IV—Exemplars

You might think that some of our property-ascribing terms—most notably, natural-kind terms like 'water' and 'gold' and 'tiger'—are defined in terms of exemplars. E.g., you might think that the proper way to define 'tiger' is to point at a bunch of tigers and to make some stipulation about the word 'tiger'. And you might think that 'wrong' should be defined in some such way as well. E.g., you might think that the proper way to define 'wrong' is to point at some specific bunch of actions—e.g., a bunch of actions that includes actions in which a person is pushed off of a bridge to stop a trolley but doesn't include actions in which a lever is pulled to switch a trolley onto a different track—and to say that 'wrong' expresses the wrong-like property (or perhaps the most natural wrong-like property) that applies to precisely *that* bunch of actions.

Given these remarks, you might think that facts about *what the wrong actions have in common* are relevant to the what-is-wrongness question. But this is confused. If we help ourselves (just for a moment, and just for the sake of simplifying things) to Lewisian realism about possibilia, we can bring this point out very clearly. The problem is that for every answer to the what-is-wrongness question—i.e., for every wrong-like property—there's a set of actions of the relevant kind, i.e., a set of (actual and possible) actions that have the given wrong-like property in common.[9] Consider, e.g., a dispute between a Millian and a Kantian about what wrongness is. It's completely unhelpful to say that this dispute is settled by facts about what the wrong actions have in common because (a) there's a set of (actual and possible) Mill-wrong actions (and what they have in common is being Mill-wrong); and (b) there's a set of (actual and possible) Kant-wrong actions (and what they have in common is being Kant-wrong); and (c) there's not an independent fact of the matter about which of these sets (if either) is the set of *wrong* actions. Or to put point (c) slightly differently: questions like 'What do the wrong actions have in common?' and 'Which set of (actual and possible) actions is the set of *wrong* actions?' are essentially *equivalent* to—or, better, they're *settled by the same facts as*—the what-is-wrongness question. So, if you like, you can say that the answer to the what-is-wrongness question is settled by the answer to the what-do-the-wrong-actions-have-in-common question (or the which-actions-are-the-*wrong*-actions question); but (i) the opposite claim—that the answers to the latter two questions are settled by the answer to the what-is-wrongness question—actually seems more apt (because the what-is-wrongness question seems to be the most basic of these questions); and more importantly in the present context, (ii) we can't make any progress by moving from the what-is-wrongness question to the what-do-the-wrong-actions-have-in-common question (or the which-actions-

[9] Of course, if an analysis picks out a property that *couldn't* be instantiated, then the relevant set will be the *empty* set—i.e., there won't be any (actual or possible) actions that instantiate the given property—and you might think that facts like this could be relevant to the what-is-wrongness question. I'll consider this suggestion below, in Sect. 15.3.8.

are-the-*wrong*-actions question) because there aren't independent facts that settle the latter two questions. On the contrary, these questions are all settled together, by the same facts.

15.3.6 Moral Privilege

You might object that while I've been taking the central question to be a *semantic* question, it's actually a *moral* question; in other words, you might think the question isn't what we *do* mean by 'wrong' but what we *ought* to mean by 'wrong'. But there's an obvious problem with this proposal—moral theories are package deals. Put differently, the problem is that for each wrong-like property, there's a corresponding *ought*. So, e.g., we ought_1 to use 'wrong' to express $\textit{wrongness}_1$; and we ought_2 to use 'wrong' to express $\textit{wrongness}_2$; and so on. So the question now becomes: 'Which of these ought-like things is the *real* ought?' And the meta-ought-question becomes: 'What sorts of facts determine the answer to the what's-the-real-ought question?' But this is exactly analogous to the situation we were in before, and so no progress has been made.[10]

15.3.7 Theoretical Role

Some people think that when we ask questions like 'What is free will?', 'What is knowledge?', and so on, one desideratum for an adequate answer is that the concept (or property, or whatever) that we zero in on has to be capable of doing the work that it's supposed to do in our best theory. But in the present case, this is no help. The problem is that for each wrong-like property, there's a corresponding moral theory. Theory_1 says that an action is wrong iff it's wrong_1, so if this theory is true then *wrongness* is $\textit{wrongness}_1$; and theory_2 says that an action is wrong iff it's wrong_2, so if this theory is true then *wrongness* is $\textit{wrongness}_2$; and so on. But now the question becomes: 'Which of these theories is true?' And the meta-theoretical-question becomes: 'What sorts of facts determine which of these theories is true?' And so, again, no progress has been made.

[10]A related view, suggested by Eklund (2017), is that the word 'wrong' has a certain *normative role*, and this role fixes the reference of '*wrongness*'. But this just seems to push the problem back a step. For if normative role really determines reference (and that's a big *if*, for it seems that all of the wrong-like properties could be employed in normative ways), then it would seem that there are *many* normative roles (or normative-like roles), and we can ask what determines which of these roles is the role of 'wrong', and so we'll be right back where we started.

15.3.8 Coherence and Arbitrariness

You might think that when we ask conceptual-analysis questions like ‘What is free will?’ and ‘What is *wrongness*?’, we’re not just trying to report what the ordinary-language meanings of the relevant expressions are; we’re also trying to, in some sense, *clean up* ordinary usage—by, e.g., eliminating incoherence and arbitrariness. I don’t think this is true—and I’ve argued as much in my (2016)—but that doesn’t matter here. For even if we grant that facts about coherence and arbitrariness are relevant to determining what *wrongness* is, this won’t change anything important about the present dialectic. This is because there are coherent/non-arbitrary versions of *all* of the wrong-like properties we might be concerned with here. In other words, there’s a vast plurality of *coherent and non-arbitrary* wrong-like properties. And we’ll still need to say what determines which of these properties counts as *wrongness*. And we’ll still be left with the view that this is determined by facts about us.

15.4 Pushing the Argument Further

So far I’ve argued that (a) if there are any wrong-like properties, then there’s a vast plurality of them, and (b) which of these properties counts as *wrongness* is determined by facts about folk meaning. But in arguing for these claims, I assumed that (i) we use ‘wrong’ in a property-ascribing way, so that it’s *supposed to* express a property, and (ii) our usage and intentions concerning ‘wrong’ zero in on a *unique* property, as opposed to no property or many properties. I now want to drop these two assumptions and make the following two claims:

(C) Whether we use ‘wrong’ in a property-ascribing way depends on facts about *us*—about our usage, intentions, and practices concerning ‘wrong’.

(D) Assuming that there are wrong-like properties (and that we use ‘wrong’ in a property-ascribing way), whether our usage and intentions concerning ‘wrong’ zero in on a unique property, or many properties, or no property at all, is determined by facts about us.

I don’t have the space to argue for (C) and (D) here, but the arguments for these two claims are deeply analogous to the section-15.3 argument for thesis (B), and I think it’s pretty obvious that if the latter argument is cogent, then the former arguments are cogent as well. Moreover, it’s worth noting that (C) and (D) are both fairly obvious—much more obvious than (B). Indeed, (C) strikes me as more or less trivial. (D) is perhaps a bit less obvious than (C), so let me just make two quick points about (D)—one about the possibility that ‘wrong’ doesn’t express any property at all, and one about the possibility that it expresses many properties.

First point: as long as we don't use 'wrong' to express a property that *couldn't* exist—e.g., a property that both is and isn't identical to *Kant-wrongness*[11]—then it expresses at least one property (assuming that there *are* properties and that we use 'wrong' in a property-ascribing way).

Second point: whether 'wrong' picks out a unique property or many properties depends on whether our usage and intentions here are perfectly precise. If they are, then 'wrong' picks out a unique property; if not, it picks out many properties—in particular, all the properties that are consistent with our usage, intentions, practices, and so on.[12] (You might think that even if we're not perfectly precise, other facts—e.g., facts about joint-carvingness—could come in to provide a unique referent for '*wrongness*'; but I already argued against this suggestion in connection with thesis (B).)

15.5 Deflationary Consequences for Normative and Metaethics

Let *moral folkism* be the conjunction of (A)-(D) together with the claim that similar theses hold for other moral properties, e.g., *moral goodness*. I obviously haven't given a complete argument for moral folkism, but I think I've said enough to make it seem plausible. I now want to discuss what follows from this view.

The first point I want to make is that if moral folkism is true, then many normative and metaethical debates are settled by facts about folk meaning—i.e., by empirical facts about what we mean by our words. For example, (i) the question of whether some non-cognitivist or expressivist view is true is settled by facts about folk meaning; and (ii) questions about the nature of *wrongness*—e.g., whether it's a natural or non-natural property, and whether it's a normative or descriptive property—are settled by facts about folk meaning; and (iii) normative ethical disputes about what the right moral system is—whether it's Kantian or utilitarian or whatever—are settled by facts about folk meaning.[13] If moral folkism is true, then there's nothing about the nature of objective non-linguistic reality at issue in connection with any of these debates; they're all settled by empirical facts about the heads of the folk.

---

[11]Note that this is different from scenarios in which we use 'wrong' to express a property that *couldn't be instantiated*. According to PP, properties like that *do* exist.

[12]Suppose that *wrongness*$_1$ and *wrongness*$_2$ both fit with our usage and intentions concerning 'wrong' and that type-T actions are wrong$_1$ but not wrong$_2$. Then on my view, there's no fact of the matter whether type-T actions are wrong. This is exactly analogous to what happens when mathematical and physical predicates are imprecise.

[13]This, at any rate, is true if we interpret this debate as being about the nature of moral properties like *wrongness* and *goodness* and so on. If we interpret the debate as being out the *extensions* of moral predicates, then it's *not* settled by facts about meaning. In this case, the debate is analogous to *applied*-ethical debates—which I'll discuss in Sect. 15.6.

Note, however, that if we use ‘wrong’ in a property-ascribing way, then the question of whether moral realism is true is *not* settled by facts about folk meaning; for if ‘wrong’ expresses some specific property, then the question of whether that property is instantiated (and, hence, whether moral realism or error theory is true) is determined by objective facts about reality, not by facts about us.

15.6 Why Moral Folkism *Doesn’t* Entail That *Applied*-Ethical Disputes Are Settled by Facts About Folk Meaning

Consider an ordinary dispute between two people—two ordinary members of our culture—about whether actions of some kind K (e.g., meat-eating actions, or whatever) are wrong. Moral folkism does *not* imply that disputes like this are settled by facts about folk meaning. On the contrary, if we use ‘wrong’ in a property-ascribing way, then on the view I’m putting forward, applied-ethical disputes like this are settled by objective facts about the nature of kind-K actions—about whether these actions instantiate the property of *wrongness*, i.e., the property expressed by the English word ‘wrong’.

Now, if we *don’t* use ‘wrong’ in a property ascribing way, then applied-ethical disputes might be misguided in some way; but this is irrelevant to the point I’m making here—that *moral folkism doesn’t imply* that applied-ethical disputes are settled by facts about folk meaning.

Also, even if we use ‘wrong’ in a property-ascribing way, it’s unlikely that our usage and intentions are precise enough to zero in on a *unique* wrong-like property, and given this, it could be that some applied-ethical disputes are indeterminate—i.e., there could be no fact of the matter whether actions of the relevant kind are wrong because they count as wrong on some legitimate precisifications of ‘wrong’ but not others (for more on this, see Sect. 15.4). But this doesn’t undermine the claim I’m making here—that ordinary disputes about whether kind-K actions are wrong are settled by facts about the nature of kind-K actions and not by facts about meaning.

It’s important to remember in this connection that ordinary people don’t usually endorse theories of what *wrongness* is; and when they argue about whether things like eating meat are wrong, they’re not usually in agreement that eating meat is wrong$_1$ but not wrong$_2$, so that what they’re “*really* debating”, in some sense, is whether *wrongness* is identical to *wrongness*$_1$ or *wrongness*$_2$; they’re just arguing about whether eating meat is *wrong*—period.

(It may be that *some* applied-ethical disputes are merely verbal. For it may be that (a) some people use moral terms in idiosyncratic ways and/or (b) there are sub-communities within our culture in which words like ‘wrong’ express different properties. My own view is that this is less common than you might think and that *most* applied-ethical disputes are *not* merely verbal. But I can’t argue for this here, and it’s not relevant to the point I’m making in this section.)

15.7 Why Moral Folkism Doesn't Undermine Moral Realism (or Error Theory)

Consider the following view:

> *Moral Pluralism*: Moral folkism is true, and we use moral predicates like 'wrong' and 'good' in property-ascribing ways. Thus, on this view, there's a vast plurality of moral systems (e.g., Kantianism, Millianism, and so on), and which of these systems is true is determined by facts about us—about what properties we have in mind when we talk about morality.

If the arguments of this paper are correct, so that moral folkism is true, then moral realists (and error theorists) are committed to moral pluralism. But you might think we have good reason to reject moral pluralism, and so you might think that if moral folkism is true, then moral realism (and error theory) are false. But I'm not so sure that we have good reason to reject moral pluralism. In this section, I'll consider a few arguments against pluralism and argue that none of them is cogent.

One might argue against moral pluralism by claiming that (a) pluralism implies that all moral disputes are either merely verbal or settled by facts about meaning, and (b) this isn't true.[14] But as we already saw in Sect. 15.6, pluralism is perfectly compatible with the claim that ordinary applied-ethical disputes are *not* merely verbal or settled by facts about meaning.

But anti-pluralists don't need to claim that pluralism implies that *all* moral disputes are merely verbal or settled by facts about meaning. All they need to do, in order to undermine pluralism, is locate a *single* moral dispute—call it "MD"—that satisfies the following two conditions:

(I) Moral pluralism implies that MD is merely verbal or settled by facts about folk meaning.
(II) We have good reason to think that MD is *not* merely verbal or settled by facts about folk meaning.

Here are three different disputes that one might think satisfy these two conditions:

The twin dispute: Suppose that (a) we use 'wrong' to express *wrongness*$_1$; and (b) there's a community of Twin Earthlings who use 'wrong' to express *wrongness*$_2$; and (c) eating meat is wrong$_1$ but not wrong$_2$. Now suppose that we get into a dispute with the Twin Earthlings about whether eating meat is wrong (we say it's wrong, and they say it isn't); call this "the twin dispute".

The what-is-wrongness dispute: This is just the dispute about the what-is-wrongness question—i.e., the question of which wrong-like property counts as *wrongness*.

The dispute in which we know all the non-meaning facts: Suppose we're wondering whether actions of some kind K are wrong, and suppose that for every wrong-like property *wrongness*$_i$, we (somehow) know whether kind-K actions are wrong$_i$—e.g., we know whether they're Kant-wrong, whether they're Mill-wrong, and so on. Now suppose that

[14] Clarke-Doane (forthcoming) puts forward an argument like this, but his argument is a bit different; his point isn't that moral pluralism is false; it's that questions about moral facts aren't the important questions surrounding deliberation.

in this scenario we get into a dispute about whether kind-K actions are wrong; call this "the dispute in which we know all the non-meaning facts".

Each of these disputes gives us a different argument against moral pluralism—an argument that's generated by taking (I) and (II) to be about the dispute in question. I want to argue that none of these arguments is good. I'll start with the argument that's about the dispute in which we know all the non-meaning facts—i.e., the argument based on the following two claims:

(M) Moral pluralism implies that the dispute in which we know all the non-meaning facts is settled by facts about folk meaning.
(MM) We have good reason to think that the dispute in which we know all the non-meaning facts is *not* settled by facts about folk meaning.

I want to argue that (M) is false. To get at the central issue here, let's distinguish two different kinds of settling:

A fact F *metaphysically settles* a question Q iff it's the truthmaker of the correct answer to Q, or F makes it the case that that answer is correct, or some such thing.
A fact F *epistemically settles* a question Q *for agent A* iff, given A's epistemic situation, A can figure out the answer to Q by discovering F—or some such thing.

According to moral pluralists, the question of whether kind-K actions are wrong is not *metaphysically* settled by facts about what we mean by 'wrong'. Rather, it's settled by objective facts about whether kind-K actions actually possess the property of *wrongness*—i.e., the property expressed by the English word 'wrong'. Now, in some very weird situations (e.g., situations in which we somehow know all the non-meaning facts), this question is *epistemically* settled—*for us*—by facts about folk meaning; in other words, given what we know in this situation, we can discover whether kind-K actions are wrong by determining which property is expressed by 'wrong'. But it doesn't follow that that question is *metaphysically* settled by facts about folk meaning, and in fact, it *isn't*.

We can put the point here as follows. If we read the word 'settling' as meaning *metaphysical settling*—and I take it that this is the reading that's needed for the (M)-(MM) argument to be even initially promising—then (M) is false; moral pluralists don't have to say (and *shouldn't* say) that the dispute in which we know all the non-meaning facts is metaphysically settled by facts about folk meaning. (Also, if we read 'settling' as meaning *epistemic settling*, then I think it's pretty easy to argue that (MM) is false; but I won't bother with this here because I don't think many anti-pluralists would endorse the epistemic-settling version of the argument.)

Notice how different this is from the what-is-wrongness dispute. On the view I've argued for here, that dispute is *metaphysically* settled by facts about us. If it's built into our usage and intentions and practices that 'wrong' expresses $wrongness_{17}$, then that *makes it the case* that 'wrong' *does* expresses $wrongness_{17}$. Before we came along, $wrongness_{17}$ wasn't singled out as special in any way; it was just sitting there in platonic heaven, alongside the other wrong-like properties. It wasn't until we came along and started focusing on this property that it became correct to say that it was the property of *wrongness*.[15]

This leads us naturally to the anti-pluralist argument based on the what-is-wrongness dispute; i.e., it leads us to the following argument:

(W) Moral pluralism implies that the what-is-wrongness dispute is settled by facts about folk meaning.
(WW) We have good reason to think that the what-is-wrongness dispute is *not* settled by facts about folk meaning.

I want to discuss this argument together with the argument based on the twin dispute—i.e., this argument:

(T) Moral pluralism implies that the twin dispute is merely verbal.
(TT) We have good reason to think that the twin dispute is *not* merely verbal.[16]

I think that (W) and (T) are both true. But I think that (WW) and (TT) are both false. Indeed, I think we have good reason to think that (i) the twin dispute *is* merely verbal, and (ii) the what-is-wrongness dispute *is* settled by facts about folk meaning. I've already argued for claim (ii)—the arguments of Sect. 15.3 suggest that we should *all* say that the what-is-wrongness dispute is settled by facts about folk meaning, regardless of whether we endorse moral pluralism. And claim (i) strikes me as more or less trivial. To appreciate this, just look at what's built into the description of the twin dispute; we're supposed to assume that (a) we use 'wrong' to express *wrongness*$_1$, and (b) the Twin Earthlings use 'wrong' to express *wrongness*$_2$; and (c) eating meat is wrong$_1$ but not wrong$_2$. It seems altogether obvious that if all of this is true, then the twin dispute *is* merely verbal. Of course it is—this is just what a verbal dispute *is*. (If we don't assume that (a)–(c) are all true—if we just assume that we're in a dispute with a community of Twin Earthlings about whether eating meat is wrong—then it won't follow that the dispute is merely verbal. But in this case, moral pluralism won't imply that the dispute is merely verbal, and so the anti-pluralist argument won't go through for that reason.)

But why might one think that (WW) and (TT) are true? What "good reason" might one think we have to believe that the twin dispute is *not* merely verbal and the what-is-wrongness dispute is *not* settled by facts about meaning? The only response to this I can think of is that we have an *intuition* that the twin dispute isn't merely verbal and the what-is-wrongness dispute isn't settled by facts about meaning. (This, I think, is the driving idea behind the argument in Horgan and Timmons (1991); they claim that we have an intuition that the twin dispute isn't verbal and that this undermines certain kinds of realism.)

[15]We have to be careful how we put this point. If the question at issue is 'What is the decompositional structure of *wrongness*?', then facts about us don't metaphysically settle the question in the sense at issue here. But as we saw in Sect. 15.3.1, once we've answered the question "Which wrong-like property is expressed by 'wrong'?"—which *is* metaphysically settled by facts about us—there's nothing left to discover.

[16]This is essentially equivalent to the argument in Horgan and Timmons (1991)—although they were arguing against Cornell realism, not moral pluralism.

I don't want to deny that some people—perhaps even most people—have these intuitions. But so what?; if the arguments of this paper are right, then these intuitions are just mistaken. Moreover, I think it can be argued that we shouldn't trust intuitions of this kind anyway. Indeed, I think some people would say that the knee-jerk beliefs we're talking about here—i.e., the belief that the twin dispute isn't merely verbal and the belief that the what-is-wrongness dispute isn't settled by facts about folk meaning—aren't *intuitions* at all. But the way I want to put the point is as follows: regardless of whether these knee-jerk beliefs count as "intuitions", they're not intuitions of the kind that we should trust and take as data points in our reasoning. I can't argue for this point in depth, but I'd like to say a few words about it. It seems to me that the kinds of intuitions that we should trust, and that we should take as data points, are intuitions about the applicability and non-applicability of our concepts in real and imagine scenarios. The reason we should trust intuitions of this kind is that we have a story to tell about why they're reliable—it's because we're competent users of the relevant predicates, and the intuitions in question here are just judgments about how to use these predicates. But "intuitions" about whether the twin dispute is merely verbal, and whether the what-is-wrongness dispute is settled by facts about meaning, aren't like this at all, and we don't have any account of why these intuitions are reliable. Indeed, appealing to "intuitions" of this kind seems every bit as illegitimate as appealing to the "intuition" that moral realism is false—these just aren't the kinds of things we can know by intuition. But I can't say any more to motivate this view here.

You might respond to all of this as follows:

> You've missed the point of the twin dispute. With non-moral disputes, once we realize that a dispute is merely verbal, we stop arguing; but with the twin dispute, we don't—we still feel that there's an important dispute to be had.

My main response to this is that it doesn't matter whether we "feel" that there's an important dispute to be had with the Twins; for, to repeat, we have no reason to treat this feeling as providing good evidence for the claim that the Twin dispute isn't merely verbal. But there's a second point worth making here. Moral pluralists don't have to say that there's "no important dispute to be had" with the Twins. All they have to say is that the *twin* dispute—i.e., the dispute between us and the Twins about the specific issue of whether eating meat is wrong—is merely verbal. But (a) pluralists can claim that we could have *other* disputes with the Twins (e.g., about which moral system is pragmatically (i.e., non-morally) better, or about the pragmatic (i.e., non-factual) question of *what to do*)[17]; and (b) pluralists could appeal to recent work on metalinguistic negotiation[18] to argue that even if the twin dispute is merely verbal (and even if the what-is-wrongness dispute is settled by facts about meaning), they could still be *important*.

[17] Clarke-Doane's (forthcoming) position is that non-factual pragmatic questions of this kind (and not moral questions about what we *ought* to do) are the really important questions. And Gibbard (2003) thinks that questions about what we *ought* to do just *are* questions about what to do.

[18] See, e.g., Plunkett and Sundell (2013), Thomasson (2016), and Belleri (2017).

15.8 Is This *Real* Realism?

You might claim that since pluralistic moral realism implies that the twin dispute is merely verbal and the what-is-wrongness dispute is settled by facts about meaning, it's not a genuine version of *realism* at all—or at any rate, it doesn't give us what we *wanted* out of moral realism. (Claims in this general vicinity have been made by Horgan and Timmons (1996), Street (2006), Eklund (2017), Clarke-Doane (forthcoming), and many others.)

One response to this is to point out that the sort of realism that I'm talking about here—roughly, the view that there are some actions (or people or whatever) that possess the property of *wrongness* (or *goodness* or whatever)—is exactly analogous to ordinary kinds of realism about things like electrons and planets. E.g., realists about planets have to say that (a) which planet-like property counts as *planethood* is determined by facts about what we have in mind when we use the term 'planet', and (b) if Twin Earthlings used 'planet' slightly differently from the way we use it, then we could have a verbal dispute with them about whether Pluto is a planet.

But you might counter this by claiming that *moral* realism is different from other kinds of realism; you might claim that unlike planet realism, moral realism is *supposed to* deliver the result that disputes like the twin dispute and the what-is-wrongness dispute are *not* merely verbal or settled by facts about meaning.

Whether this is true depends on what 'moral realism' means, and it's not clear why this should matter. I think we should just distinguish two different kinds of moral realism. We can define *weak moral realism* as the view that some actions (or people or whatever) possess some moral properties (e.g., *wrongness*, *goodness*, etc.); and we can define *strong moral realism* as weak moral realism plus some extra thesis—e.g., that there are objective facts (independent of us) that make moral properties like *wrongness* and *goodness* stand out from other moral-like properties (i.e., wrong-like properties and good-like properties and so on) as special, or privileged (perhaps because *wrongness* and *goodness* are "glowing", or because there are supernatural moral joints in reality). Given this, we can say that pluralistic realism is a version of weak realism but not strong realism and leave it at that.

For whatever it's worth, I think there are multiple problems with strong moral realism. I've already pointed out that views of this kind are problematic in metaphysical, epistemological, and moral ways. But I also think that (i) strong realism fails to deliver the things that strong realists *want* (i.e., the things that we don't get from weak realism), and (ii) these extra things aren't actually desirable (i.e., weak realism already gives us everything we should want out of moral realism). But I can't argue for these claims here.

Finally, I'd like to emphasize that I'm not claiming here that weak moral realism is true. For all that I've argued, it could be that (a) we don't use 'wrong' in a property-ascribing way, so that some sort of non-cognitivism or expressivism is true, or (b) we use 'wrong' to pick out a property that isn't instantiated, so that error theory is true. Both of these views are compatible with moral folkism.

References

Balaguer, M. (1998). *Platonism and anti-Platonism in mathematics*. New York: Oxford University Press.

Balaguer, M. (2009). Fictionalism, theft, and the story of mathematics. *Philosophia Mathematica, 17*, 131–162.

Balaguer, M. (2016). Conceptual analysis and X-PHI. *Synthese, 193*, 2367–2388.

Belleri, D. (2017). Verbalism and metalinguistic negotiation in ontological disputes. *Philosophical Studies, 174*, 2211–2226.

Boyd, R. (1988). How to be a moral realist. In G. Sayre-McCord (Ed.), *Essays on moral realism* (pp. 181–228). Ithaca: Cornell University Press.

Brink, D. O. (1989). *Moral realism and the foundations of ethics*. Cambridge: Cambridge University Press.

Clarke-Doane, J. (forthcoming). *Morality and mathematics*. Oxford: Oxford University Press.

Dasgupta, S. (2017). Normative non-naturalism and the problem of authority. *Proceedings of the Aristotelian Society, 117.3*, 297–319.

Dunaway, B., & McPherson, T. (2016). Reference magnetism as a solution to the moral twin earth problem. *Ergo, 3*, 639–679.

Eklund, M. (2017). *Choosing normative concepts*. Oxford: Oxford University Press.

Gibbard, A. (2003). *Thinking how to live*. Cambridge, MA: Harvard University Press.

Horgan, T., & Timmons, M. (1991). New wave moral realism meets moral twin earth. *Journal of Philosophical Research, 16*, 447–465.

Horgan, T., & Timmons, M. (1996). From moral realism to moral relativism in one easy step. *Critica, 28*, 3–39.

Jackson, F. (1998). *From metaphysics to ethics*. Oxford: Oxford University Press.

Lewis, D. (1986). *On the plurality of worlds*. Oxford: Blackwell Publishing.

McDaniel, K. (2017). Normative accounts of fundamentality. *Philosophical Issues, 27*, 167–183.

Plunkett, D., & Sundell, T. (2013). Disagreement and the semantics of normative and evaluative terms. *Philosophers' Imprint, 13*, 1–37.

Schroeter, L., & Schroeter, F. (2013). Normative realism: Coreference without convergence? *Philosopher's Imprint, 13*, 1–24.

Sider, T. (2011). *Writing the book of the world*. Oxford: Oxford University Press.

Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies, 127*, 109–166.

Thomasson, A. (2016). Metaphysical disputes and metalinguistic negotiation. *Analytic Philosophy, 57*, 1–28.

Williams. (2018). Normative reference magnets. *Philosophical Review, 127*, 41–71.

Chapter 16
Methodology in the Ontology of Artworks: Exploring Hermeneutic Fictionalism

Elisa Caldarola

Abstract There is growing debate about what is the correct methodology for research in the ontology of artworks. In the first part of this essay, I introduce my view: I argue that semantic descriptivism is a semantic approach that has an impact on meta-ontological views and can be linked with a hermeneutic fictionalist proposal on the meta-ontology of artworks such as works of music. In the second part, I offer a synthetic presentation of the four main positive meta-ontological views that have been defended in philosophical literature about artworks and of some criticisms that can be lodged against them: Amie Thomasson's global descriptivism, Andrew Kania's local descriptivism, Julian Dodd's folk-theoretic modesty, and David Davies' rational accountability view. In the conclusion, I show the advantages of my view.

Keywords Semantic descriptivism · Hermeneutic fictionalism · Meta-ontology of music · Stephen Yablo · Amie Thomasson · Andrew Kania · David Davies · Julian Dodd

There is growing debate about what is the correct methodology for research in the ontology of artworks. In the first part of this essay, I introduce my view: I argue that *semantic descriptivism* is a semantic approach that has an impact on meta-ontological views and can be linked with a *hermeneutic fictionalist* proposal on the meta-ontology of artworks such as works of music.[1] In the second part, I offer a synthetic presentation of the four main positive meta-ontological views that have been defended in philosophical literature about artworks and of some criticisms

[1]Similar considerations apply also to other kinds of works, such as works of performance art and of installation art, whose metaphysics and ontology is often discussed by drawing analogies with works of music (see e.g. Irvin 2012).

E. Caldarola (✉)
University of Padova, Padua, Italy

J. L. Falguera, C. Martínez-Vidal (eds.), *Abstract Objects*, Synthese Library 422,
https://doi.org/10.1007/978-3-030-38242-1_16

that can be lodged against them: Amie Thomasson's *global descriptivism*, Andrew Kania's *local descriptivism*, Julian Dodd's *folk-theoretic modesty*, and David Davies' *rational accountability view*.[2] In the conclusion, I show the advantages of my view.

16.1 Semantic Descriptivism and Hermeneutic Fictionalism

In the first part of this section, I shall introduce *semantic descriptivism*; in the second part, I shall sketch out a *hermeneutic fictionalist* approach to the ontology of artworks such as works of music – showing how the embracement of that approach is linked with the adoption of a descriptivist view on the semantics of folk art-ontological terms.

The nature of things can be opaque to us. For instance, the fact that whales are mammals is a relatively recent discovery. For a long time, speakers uttering sentences like

(1) I have seen a whale

were not in the position to know that (1) implies that the speaker has seen a mammal. Still, this did not make it less true that whales are mammals. This is because the *referent* of a term like 'whale' is not determined by the ordinary speaker's conception of what a whale is (see Putnam 1975). What about the *ontological commitments* of speakers, i.e. their beliefs about what exists and what doesn't? *Semantic descriptivism* is the view, explored by Stephen Yablo in a number of writings (e.g. Yablo 2009, 2010), that we should believe only the real (or *asserted* – as opposed to *literal* or *conventional* or *compositional*) content of the sentences that we assert in ordinary or scientific contexts. (1) is a pretty straightforward case: if I claim to have seen a whale, then I am committed to the existence of whales. Consider, however, the case of mathematical objects like numbers and sets: considerations going back to Frege (1884) (see also Burgess 2008) convinced many that, if numbers exist, they are abstract objects. This view is not refuted by the fact that even speakers who don't believe in the existence of abstract mathematical objects sometimes assert sentences like

(2) The number of homeless people is growing.

Similarly, we consider angels supernatural beings, and many doubt their existence, just as many are uncertain about the existence of the Freudian superego. When ordinary speakers assert sentences like

[2] On the other hand, Aaron Ridley has argued that "a serious philosophical engagement with music is orthogonal to, and may well in fact be impeded by, the pursuit of ontological issues, and, in particular, that any attempt to specify the conditions of a work's identity must, from the perspective of musical aesthetics, be absolutely worthless" (Ridley 2003: 203. For a reply see Kania 2008a).

(3) Nixon had a stunted superego.[3]

or

(4) She's an angel

we don't normally consider the fact that (3) and (4) can be asserted by ordinary speakers (who don't believe in the existence of angels or superegos) as a ground for revising the view that angels are supernatural beings or that superegos are metaphysically problematic. Rather, we consider the fact that (3) and (4) can be asserted by speakers who do not believe in the existence of supernatural beings or in Freud's doctrines as evidence that what is *asserted* by ordinary utterances of (3) and (4) is different from the *literal* content of those sentences.

Semantic descriptivism is in line with the Quinean methodology in ontology, according to which we should make room in our ontology for those and only those entities our best scientific theories need to refer to in order to be true (see Quine 1948), but adds that we should only believe in the existence of the *Fs* when the existence of the *Fs* is a consequence of the real, asserted content of certain sentences we assert, not when it is merely a consequence of their literal meaning.

Yablo makes some more specific remarks about the *impact* of semantic descriptivism on ontology, observing that, on certain occasions, speakers negatively determine the subject matter of their discourses, i.e. they *exclude* something from the domain of their discourses (see e.g. Yablo 2009: 519–522, 2010: 3–6). This is different from a *failure* to recognize that what we are talking about includes a certain thing. For instance, it is one thing if we have evidence that speakers *fail to recognize* that what they say when engaged in number talk entails that numbers exist, while it is another thing if – as it is usually the case when someone utters a sentence like (2) – we have evidence that speakers *exclude* numbers from the domain of their discourse. In the latter case, speakers are interested in talking about homeless people, not about numbers, so they exclude numbers from the domain of their discourse. According to Yablo, acts of exclusion like the one just described *have ontological weight*: when speakers exclude the existence of numbers from the topic they are addressing by using certain sentences, *they remain noncommittal as to the existence* of numbers.

Is there evidence for the fact that, e.g. when we engage in folk number talk, we exclude numbers from the semantic content of our theory? One way for Yablo to show that there is, indeed, evidence, is pointing out that in such circumstances we are *impatient* to talk about what we want to talk about, which is not numbers, but e.g. homeless people, and we are also *indifferent* to whether numbers exist or not (see Yablo 2001: 89). We are *impatient* because when we say e.g. that the number of homeless people is increasing, we want to describe the current state of the homeless population without waiting to ascertain whether numbers exist or not, since we are aware that the number of homeless people is increasing even if it is not clear whether numbers exist or not. Furthermore, we are *indifferent* because, even if we were to

[3] See Yablo 2001: 102.

discover that numbers do not exist, it would not seem less true to us that the number of homeless people is increasing: the question of the existence or non-existence of numbers is *irrelevant* to our topic.

Let us now turn to works of music: I shall argue that semantic descriptivism has an impact on our views about the ontology of musical works. Consider an ordinary speaker (by which I mean anyone who is not an expert in musical ontology and metaphysics) who claims:

(5) I have listened to Ani DiFranco singing *Swim* live more than ten times.

What about the *reference* of (5)? There is open debate among philosophers as to what kind of objects works of music such as songs are: in particular, are they material or abstract objects?[4] On the one hand, if metaphysical debate were to establish that songs are not abstract objects, then the reference of (5) wouldn't be abstract objects. On the other hand, if metaphysical debate were to establish that songs are abstract objects, then the reference of (5) would be abstract objects. Notice that the question of the *metaphysical* status of objects should be distinguished from the question of their *ontological* status: one could, for instance, maintain that works of music should be understood as abstract objects, metaphysically speaking, although abstract objects don't exist, from an ontological viewpoint.

What about the *asserted content* of (5), as uttered by an ordinary speaker? Is it about the *existence* of a musical work? My view is that we should answer in the negative. When uttering (5), ordinary speakers, who are not experts in ontology, are not interested in asserting the ontological view that musical works exist. Ordinary speakers are, instead, asserting that they have listened to certain musical performances that stand in some particular relation with each other, although they don't know what exactly that relation is. Talking of such performances as of performances *of the same musical work* is an effective way to imply that there is such a relation, without having to spell out what it is exactly and, relatedly, without taking sides on the debate over whether musical works *exist* as objects that are distinct from any of such performances.

My view is that there is an analogy between folk number talk and ordinary talk about musical works: as we have seen, according to Yablo, when we say that the number of homeless people is increasing we want to say something without waiting for ascertaining whether numbers exist or not, because we are aware that the number of homeless people is increasing even if it is not clear whether numbers exist or not; similarly, when we talk about Ani DiFranco's *Swim* we need a way to talk about certain performances standing in some particular relation with each other, without waiting for ascertaining whether musical works exist or not.

Consider also a second analogy: according to Yablo, even if we were to discover that numbers do not exist, it would not seem less true to us that the number of homeless people is increasing, because, when we say things such as

[4] E.g. Dodd (2007) defends the view that they are abstract objects, while Caplan and Matheson (2006) argue that they are collections of concrete particulars such as scores and performances.

(2) The number of homeless people is growing

the question of the existence or non-existence of numbers is irrelevant to our topic; the same is true when we say, for instance, that Ani DiFranco's *Swim* is a more recent work than Patti Smith's *Horses*: we would not take our claim back, were we to discover that works of music do not exist.

In particular, we can distinguish between two scenarios. In scenario A, metaphysical debate has established that, if they exist, musical works are concrete objects. In such case, the *asserted content* of sentences like (5) is not about *the existence* of works of music as concretes, but the (*purported) reference* of sentences like (5) is some kind of concretes. In scenario B, metaphysicians establish that musical works are abstract objects. In such case, the *asserted content* of sentences like (5) isn't about *the existence* of works of music as abstract objects, but the (*purported) reference* is abstract objects. In what follows, I shall argue that, in both scenarios, if ontologists establish that works of music don't exist, our folk-talk about works of music needs no revision, because it can be interpreted as mere fictional talk.

The two analogies with folk number talk provide evidence for the fact that speakers exclude the existence and nature of works of music from the domain of their discourse: they remain noncommittal as to the existence of works of music and their nature. If, when ordinary speakers utter (5) in scenario A, they are not asserting that there *exists* a concrete object that is Ani Di Franco's *Swim*, how should we interpret their statements? And if, when ordinary speakers utter (5) in B-type scenarios, they are not asserting that there *exists* an abstract object that is Ani Di Franco's *Swim*, how should we interpret their statements? Here, a comparison with Yablo's view about folk number talk is again helpful.

Yablo's view (e.g. 2001) is that when we say e.g. that the number of homeless people is growing we are talking metaphorically. The literal content of our talk is that there is something, i.e. the number of homeless people, and that this thing is growing; the real (i.e. asserted) content of our talk is that there are more homeless people than there used to be. According to Yablo (e.g. 1998: 245ff.), there is a mechanism defining the real content of our number talk as a function of the literal content of our number talk. The function that links literal content and real content of our number talk can be identified in terms of what Kendall Walton (1990, 1993) has called 'principles of generation' in 'games of make-believe'.

A principle of generation is a function that links what is true in the real world to what is true in the *fictional world*. The general scheme is: '[According to the fiction F] iff G' – where G (the *prop* of the game of make-believe) stands for what is really the case, and F (the *content* of the game of make-believe) stands for what is fictionally the case. In *prop oriented* games of make-believe we take advantage of the fact that, based on our knowledge of the principles of generation, we can say things about the real world by means of saying things which are only fictionally true (see Liggins 2014: 603). For instance, when we say:

(6) Juliet is the sun

we launch a game of prop oriented make-believe which allows us to talk about Juliet's beauty (a thing about the real world) by means of saying things that are true within the make-believe world where we pretend that Juliet is the sun. Now, according to Yablo (e.g. 2001), talk about numbers is a case of metaphorical, make-believe talk that is *prop* oriented: by means of talking about make-believed objects (numbers) we manage to say things about the real world. For instance, when we say that the number of homeless people is increasing in the game of make-believe where we pretend that numbers exist, we manage to say that (in the real world) there are more homeless people than there used to be.

In general, *fictionalism* about a discourse claims that the sentences of the discourse are not true, although they are useful (see Liggins 2011). As is well known, the literature distinguishes between revolutionary and hermeneutic fictionalists: revolutionary fictionalists about the sentences of a certain realm of discourse typically recommend that we should cease to believe that such sentences are true, but we should still carry on using the sentences, accepting them without believing them.[5] What they recommend is *a change of attitude* towards the sentences at issue. Hermeneutic fictionalists, on the other hand, claim that we do not currently believe what the sentences say, even if we seem to believe them: no change of attitude towards such sentences is required (see Liggins 2011).

The embracement of a descriptivist semantics is a reason for Yablo to opt for hermeneutic fictionalism: Yablo looks at the asserted content of our number talk and does not find trace of commitment to the existence of numbers, concluding that there is no evidence that we believe the truth of our literal talk – *contra* the revolutionary fictionalist, who argues that we are wrongly committed to the existence of numbers when we talk about them. Another reason for Yablo to opt for hermeneutic fictionalism is that it provides the best explanation of certain aspects of the practice of folk number talk: David Liggins has called "'abstract expressionism' [. . .] the doctrine [held by Yablo and others] that mathematics is useful in science because it helps us to say things about concrete objects which it would otherwise be more difficult, or perhaps impossible, for us to say" (Liggins 2014: 600). Number talk, *qua* fictional talk, according to this view, does not merely tell us how we think about numbers (as the revolutionary fictionalist claims), but is a useful expressive instrument to describe how things are. For instance, we can say "The number of sheep is square" instead of uttering the much more complicated sentence "The number of sheep is zero, or the number of sheep is one, or the number of sheep is four, or the number of sheep is nine . . . " (an infinite disjunction) (see Liggins 2014: 604, commenting on Yablo 2002: 231).

Let us now go back to the meta-ontology of works of music. I submit that we should opt for a hermeneutic fictionalist meta-ontology in this realm as well. My

[5]The distinction between revolutionary (prescriptive) and hermeneutic (descriptive) nominalism was introduced by Burgess (1983), famously used in Burgess and Rosen (1997, p. 6), and applied to fictionalist accounts in Stanley (2001).

view is that when speakers utter sentences like (5), they are engaging in fictional talk. Let me try to unpack this claim.

In scenario A, by talking about the song *Swim* as a concrete object, speakers are talking about a certain set of particular performances, and there is no evidence that they believe that the song Swim *exists* as a concrete object. In particular, they pretend to talk about the song as a concrete object in order to talk about aspects that the performances have in common (e.g. that they are all based on the same score). In scenario B, by talking about the song *Swim*, speakers are talking about an abstract object with certain physical instantiations (performances), but there is no evidence that they believe that such abstract object exists. In particular, they pretend to talk about an abstract object in order to talk about certain aspects of the song (e.g. that the *same* song seems to be executable on *different* occasions).

Abstract expressionism, I submit, is true also about our talk of musical works: the value of musical discourse about works lies in the fact that it allows us to say true things about the world in a really effective way (e.g. that sequences of notes written down by composers and their performances stand in certain relations with each other). If, for instance, I say

(7) I just listened to musical work *X* by composer *Y*

I manage to convey succinctly a nominalist belief that it would otherwise be very complicated to express, i.e. the complex nominalist history of the relations (historical, causal, normative) holding between the sequence of notes written down or sounds played by *Y* between time T_1 and T_2 and the sequence of sounds played between time T_3 and T_4 (by *Y* or by someone else), while, at the same time, succeeding in conveying the idea that performances do not stand alone, but are causally, historically, and normatively related to something else. Fictional talk of musical works, then, has representational advantages over nominalist descriptions of music.

What are the principles of generation in games of make-believe where we pretend to talk about works of music in order to talk about certain aspects of performances? What is the function that links what is true in the real world to what is true in the fictional world in this case?

As for scenario A, the fictionalist about the ontology of artworks can rely on the work of the nominalist to find the best principles of make-believe.[6] Principles of generation are simply biconditionals of a certain form – '[According to the fiction F] iff G' – and nominalists are committed to certain bi-conditionals: e.g. they claim that it is correct to call a certain collection of particulars (e.g. a score and certain performances) '*Swim*' if and only if there occur certain conditions in the concrete world (i.e. if such-and-such a score and such-and-such performances stand in certain relations with each other). The fictionalist, then, should claim that *it is fictional* that the song *Swim* exists as a concrete object if and only if there occur certain conditions in the concrete world. Instead of a function that brings us from the existence of

[6]E.g. Caplan and Matheson 2006.

something in the real world to the existence of something else in the real world, we have a function that links what is true in the real world (that such-and-such things exist and such-and-such actions take place) to what is true in the fictional world (that a certain work of music exists).

As for scenario B, my proposal is that the fictionalist about the ontology of artworks can rely on the work of the artifactualist about the ontology of works of music to find the best principles of make-believe.[7] The fictionalist just has to take the artifactualist theory that is considered most explanatory and simple and interpret its claims in fictionalist terms. Principles of generation are simply biconditionals of a certain form and artifactualists are committed to certain bi-conditionals. While the artifactualist claims that a certain abstract object *X* comes into existence if and only if there occur certain conditions in the concrete world, the fictionalist should claim that *it is fictional* that a certain abstract object *X* comes into existence if and only if there occur certain conditions in the concrete world. So, for instance, what the artifactualist describes as the steps necessary to the creation of a work of music are, for the fictionalist, what counts as props in the game of make-believe where we talk of musical works. Instead of a function that brings us from the existence of something in the real world to the existence of something else in the real world, we have a function that links what is true in the real world (that such-and-such things exist and such-and-such actions take place) to what is true in the fictional world (that works of music are created). To put it in a slogan, the principles of creation (e.g. of musical works, in the artifactualist's account) can be re-cast as principles of generation (e.g. of fictional truths about musical works) in my proposal for a fictionalist meta-ontology of works of music.

Yablo claims that his "rationale for fictionalism" in the ontology of mathematics is that "maybe, it gives the most plausible account of the practice" of uttering applied mathematical statements (Yablo 2001: 19). It seems to me that the same rationale could be given for the kind of fictionalism in the meta-ontology of art I have suggested: maybe, it gives the most plausible account of the practice of talking about works of music.

To recap and conclude, here is, in a nutshell, my meta-ontological view. When ordinary speakers utter sentences like (5), they are talking fictionally. Assuming that philosophers establish that works of music are concretes (scenario A), ordinary speakers don't need to revise their folk statements about works of music: if it turns out that works of music exist as concretes, then ordinary speakers happen to talk, fictionally, about something that is true in the real world (i.e. that works of music exist as concretes); if it turns out that works of music don't exist as concretes, then

[7] Artifactualists about artworks argue that artworks are created artifacts. Some artifactualists about works of music argue that they are created abstract artifacts. For instance, Jerrold Levinson argues that if an author selects and writes down or plays certain notes in a certain order, with the intention that they have a normative role, i.e. the role of "establishing a rule to reproduce the sounds [referred to by such notes] in a certain way following the indications of a particular, historically-situated musical mind [i.e. her own mind]" (2012: 54), then a tonal-instrumental structure is created in the real world – i.e. a work of music (a generic entity that can have instantiations) begins to exist.

ordinary speakers are merely talking fictionally about them. Similarly, assuming that philosophers establish that works of music are abstract objects (scenario B), both if they establish that works of music exist and if they establish that works of music don't exist our folk-talk about works of music needs no revision, because it can be interpreted as fictional talk about abstract objects. In case works of music exist, it is fictional talk that happens to tell the truth about the real world; in case works of music do not exist, it is mere fictional talk. In all scenarios, then, our folk statements about musical works are fictional and are not in contrast with our ontological pursuits.

In the next section, I shall briefly introduce the four main competitors to the meta-ontological view I have put forward. In the last section of the paper, I shall argue that my proposal presents some advantages.

16.2 Alternative Views

16.2.1 Amie Thomasson's Global Descriptivism

According to Thomasson, while debating the ontology of artworks we should maintain that

> the only plausible views will be those that simply make explicit the conditions for existence and identity built into our practices of treating works of art as here or there, surviving and being destroyed, etc. – it can't turn out that these practices are all wrong, and we are all terribly mistaken about what sorts of things works of art really are. [Thomasson 2006: 251-252]

Thomasson claims that, in building our views on the ontology of artworks, we should rely on what we can infer about the existence and identity of artworks from those practices to which the existence and identity of artworks are relevant (e.g. practices such as talking about the survival of artworks and locating artworks). In other words, we should discard the hypothesis that works of art have a thought-independent nature and adopt a descriptivist approach to the ontology of artworks. This is actually Thomasson's *general* approach to ontology: she endorses "descriptivism on a *global* scale" (Dodd 2013: 1054), arguing that all there is to discover about the ontological status of *any* kind of objects is just "what our practices themselves determine" (Thomasson 2005: 228).

The *first ground* of Thomasson's take on the meta-ontology of artworks is her view that the nature of any kind of objects is mind-dependent. More precisely, this is the view that, if '*K*' is a kind-term that refers to a certain kind of objects, then all we have to consider in order to tell whether something is an object of kind *K* is established by the ontological conception of *K*s that grounders of the term '*K*' hold (see Thomasson 2007b: 59; Dodd 2013: 1062). For instance, all we have to consider in order to tell whether something is a cell is the ontological conception of cell implied by grounders of the term ‘cell’.[8] The *second ground* of Thomasson’s view on the meta-ontology of artworks is her more specific view on the mind-dependency of artefacts, according to which grounders of a certain artifactual kind-term link the conceptual content of the kind-term with the intentions of the makers of the artefacts at issue so that, in order to be an artefact of a certain kind, an object has to be the product of the successful intention to create something of that kind (see Thomasson 2007b: 59; Dodd 2013: 1062). For instance, in order to be a painting, an object has to be the product of a successful intention to create a painting.

Thomasson’s meta-ontological stance has an interesting consequence for the ontology of those artworks, such as e.g. symphonies, whose conception – according to some scholars[9] – seems to include both the idea that they are abstract objects (because they cannot be identified with any of their particular performances or with their scores) and the idea that, unlike typical abstract objects (e.g. numbers), they are created and not eternal (e.g. Beethoven’s *Symphony No. 9*, which was composed by Ludwig van Beethoven between 1822 and 1824): if we apply Thomasson’s view to e.g. symphonies, it follows that we should let our conception that symphonies are created by their composers shape our ontological view of them and therefore make room, in our ontology, for abstract objects that are created, and not eternal – a quite peculiar kind of abstract objects (see Thomasson 2004, 2005, 2006).

I shall now present three criticisms that have been lodged against Thomasson: one concerning her view about reference-fixing, one concerning her solution to the so-called ‘*qua* problem’ (see note 8 below), and one concerning the view of the ontology of art that emerges from her meta-ontological position. First, as Dodd (2012, 2013) argues, Thomasson’s view of reference-fixing is not the only way out of the *qua* problem: there is an alternative view of reference-fixing that we could adopt and which does not issue in meta-ontological descriptivism. Dodd writes:

> Even if we agree that external, causal factors are insufficient to fix reference, and even if we further agree that the ontological conception of *K*s possessed by grounders plays a crucial reference fixing role, the claim that this conceptual content determines reference by acting as a template for items to fit is controversial, and thus requires argument. Indeed, this is especially true, since there are other potential explanations of how conceptual content

[8] Thomasson opts for the view of reference-fixing just described because it offers a solution to the so-called ‘*qua* problem’ (see Devitt 1981), which arises for those who hold that reference-fixing is a purely causal matter, that has no descriptive aspect: “in order to succeed in naming a certain dog ‘Spot’, I must at least know what kind of thing the nominatum-to-be is: I must at least know that he is (say) an animal. If I think he is merely an inanimate spot in my field of vision, I will not have succeeded in naming him. Now, to know what kind of object one is naming is to conceptualize that object, to think of it as an object of a certain sort, as (in other words), satisfying a certain predicate. It is thus to think of it qua such-and-such. Thus, if an act of reference-fixing is to be successful, the reference-fixer must think of the referent-to-be under a certain description – one that that object or individual actually satisfies. If this is right, however, then the event of reference-fixing cannot be conceived of in purely causal terms” (Reimer and Michaelson 2017). For Thomasson’s discussion of the *qua* problem see e.g. Thomasson (2007b).

[9] See e.g. Levinson 2012.

> determines reference that do not have the descriptivist implications of Thomasson's own. [Dodd 2013: 1063]

Dodd criticizes Thomasson's view on the mind-dependency of the nature of any kind of objects basing on the above observation, which stresses that Thomasson does not provide an argument for her view about conceptual content acting as a template for items to fit and thereby determining reference. Dodd (2012) argues that an alternative explanation of how conceptual content determines reference can be given by developing on Gareth Evans' (1973) account of reference fixing. In a nutshell, on a Evansean account, we agree with Thomasson that what determines the reference of a kind-term is the conceptual content associated with the kind-term by those responsible for introducing it in the language, but we also argue that "a general term refers to the kind of entity figuring in the sample constituting the dominant causal source of the term's associated body of information" (Dodd 2012: 84). So, instead of holding that, if '*K'* is a kind-term that refers to a certain kind of objects, then all that we have to consider in order to tell whether something is an object of kind *K* is established by the ontological conception of *K*s that those who are responsible for introducing the term '*K*' in our language hold, we hold that, if '*K*' is a kind-term that refers to a certain kind of objects, then all that we have to consider in order to tell whether something is an object of kind *K* is established by what e.g. the source of information about *Ks* that matters the most to us, or the source of information about *Ks* that dominates our view of *Ks*, tell us about what *Ks* are (see Evans 1973: 16 and Dodd 2012: note 48, p. 85). On Thomasson's view, there is no room for the possibility that the conceptual content associated with a certain kind-term by those responsible for introducing it in the language be inaccurate "with respect to the existence and identity conditions of entities of that kind" (Dodd 2012: 85), and therefore we end up admitting into our ontology 'exotic' entities such as created abstract objects. On the Evansean view, instead, "since the conceptual content associated with a kind term fixes the term's reference by causal origin, and not by fit, [. . .] this content can contain substantial inaccuracy with respect to the existence and identity conditions of things of that kind" (Dodd 2012: 85), and therefore we can e.g. choose not to admit created abstract objects into our ontology, on the ground that we judge inaccurate the view that e.g. symphonies are both created objects and abstract objects, given our best philosophical conception of abstract objects. Opting for the Evansean view on reference-fixing, then, would allow us to avoid Thomasson's descriptivist stance on meta-ontological issues, while at the same time agreeing with Thomasson that a purely causal account of reference-fixing is to be rejected (see note 8 above). Dodd concludes that Thomasson owes us an explanation of why her account of reference-fixing is better than the Evansean one (Dodd 2013: 1063).[10]

[10]In defense of the Evansean account, Dodd (a) conducts a cost-benefit analysis, comparing it to Thomasson's account and concluding that the Evansean account is superior (Dodd 2012: 85–91) and (b) replies to some objection that might be raised against it (Dodd 2012: 91–95).

While Dodd shows that Thomasson's view of reference-fixing is not the only way out of the *qua* problem, it can also be argued – as Stephen Yablo (2014: 498–500) does, developing on a remark by Paul Boghossian (1994: 119) – that Thomasson's solution to the *qua* problem is not successful, because it requires us to accept an unjustified claim. While it is acceptable to claim that in our practices we assume that e.g. the sentence 'Everest is a mountain' must *be held* true in order for the term 'Everest' to have a certain meaning, we have no reason to accept e.g. that the claim that 'Everest is a mountain' must *be* true in order for the term 'Everest' to have a certain meaning. Thomasson's solution to the *qua* problem, however, requires us to hold e.g. that 'Everest is a mountain' must *be* true in order for the term 'Everest' to have a certain meaning, given that our linguistic practice considers 'Everest is a mountain' as constitutive of the meaning of the term 'Everest'. So, since, according to Thomasson, accepting descriptivism is a consequence of adopting a view of reference-fixing that solves the *qua* problem and since, as Yablo shows, such solution to the *qua* problem is problematic, we must conclude that Thomasson hasn't provided conclusive reasons to accept descriptivism and admit into our ontology, among other things, entities such as created abstract objects.

A third criticism of Thomasson's proposal has been put forward by David Davies (2017: 125–127). According to Thomasson, Davies stresses, "metaphysical inquiry into the nature of works of art of different kinds cannot be revisionary of the conceptual content conferred upon the relevant art-kind terms by ordinary users of those terms, since it is such content that determines the kind of thing to which the term refers" (Davies 2017: 125). Davies believes that Thomasson's anti-revisionary view of the ontology of art betrays a misconception of the project guiding research in this field. According to Thomasson, art ontologists should deal in conceptual analysis, focussing on the uses of our folk-ontological art terms: this stance brings Thomasson to argue e.g. that Gregory Currie's theory that paintings are action types, rather than individual artefacts, should be rejected, because it is revisionary of the conceptual content conferred upon the kind-term 'painting' by ordinary users of the term (see Thomasson 2007a: 191; Davies 2017: 126). Davies, however, claims that Thomasson misrepresents what Currie is doing, as a consequence of her view that art ontologists should deal in conceptual analysis: what matters to Currie and other art ontologists are not folk art concepts (which they believe need not be corrected and can stay the same), but rather "the way in which the referent[s] of th[ose] folk concept[s] enter into our appreciation, [production, and criticism of artworks]" (Davies 2017: 126). Thomasson's anti-revisionary view, then, cannot do justice to the fact that "ontology of art involves [. . .] the codification of a practice in a way that clarifies the role that certain things play within it" (2017: 127) – it is implicit in Davies' argument that this is evidently the right way of conceiving of the art ontologist's endeavours.

To conclude this brief presentation, the lesson to bring home, I believe, is that the literature has shown that Thomasson's proposal presents several shortcomings and that it should be either radically revised or abandoned.

16.2.2 Andrew Kania's Local Descriptivism

While, as we have seen, Thomasson holds that we should adopt a descriptivist stance not only in the ontology of art but in all ontological pursuits, Andrew Kania (2008a, b, 2012) holds that fundamental metaphysics might or might not be descriptive and argues for descriptivism limited to the ontology of music. As Dodd (2013: 1054) observes, his view is then a form of *local descriptivism* in the ontology of art. Here's Kania expressing his methodological stance:

> though musical works must have some fundamental nature, if they exist at all, it does not follow that they will be metaphysically fundamental in the same sense as substances or tropes might be. Musical works get made in complex social situations – they are cultural artifacts. Part of what this means is that their nature and properties depend not only on individual minds but on complex interactions between many different minds, and thus they will be at least as high up on the ontological ladder as minds. [Kania 2008b: 437]

The above passage shows that Kania is, like Thomasson, committed to the view that facts about musical works are not objective, but depend on the ontological beliefs implicit in our music-related practices (e.g. appreciating music, composing music) – a view labeled by Dodd "the determination thesis" (Dodd 2013: 1055). This commitment brings Kania to embrace local descriptivism in musical ontology: "It is surely something like this thought that is behind the idea that musical ontology is ineliminably descriptive, because if this is correct, then how musical works *are* depends upon how people *think about* them" (Kania 2008b: 437–438).

To my knowledge, criticisms against Kania's meta-ontological stance have come from Dodd (2013) and Philip Letts (2015). Dodd (2013: 1059) observes, *in the first place*, that Kania (1) assumes that musical works are artefacts, and (2) that therefore they are human creations and from this concludes (C1) that, *qua* human creations, they must ontologically depend on our practices, i.e. on "complex interactions between many different minds" and (C2), that in dealing with the ontology of musical works we should thus be descriptivists. However, as Dodd stresses – mentioning a remark by Thomasson (2007b: 52) – it is not true that *since works of music are human creations* then they must ontologically depend on our practices, because there are human creations (e.g. genetically modified plants) that *exist* because they are intentionally created by human beings and that, nevertheless, *do not have a certain ontological nature* (e.g. are plants, rather than animals) because they are intentionally created by human beings.[11]

In the second place, Dodd (2013: 1060–1061) identifies a tension between Kania's view that our practices determine the ontological nature of musical works, *qua* artefacts, and Kania's view that, when it comes to fundamental metaphysics, we might have to abandon meta-ontological descriptivism. Suppose, Dodd explains,

[11] Compare also what David Liggins says about "languages or conventions": "Even if such things depend for their existence on human activity, [...] the dependence claim is not 'Whatever we believe about these things, it is true because we believe it': a sensible account of languages or conventions should allow that we are sometimes mistaken" (Liggins 2010: 75).

that our non-descriptivist meta-ontological stance in fundamental metaphysics brings us to believe that all persisting objects, artefacts included, are instantaneous temporal stages: according to this view e.g. pencils are instantaneous temporal stages. Now, this view would be at odds with the folk view we – *qua* local descriptivists in the meta-ontology of artefacts – would have to hold about the persistence of e.g. pencils, according to Kania: thinking of pencils as instantaneous temporal stages definitely doesn't reflect our folk beliefs about pencils! This example shows that "it is naive to think that the determination thesis can be maintained for artefacts whilst the thesis of metaontological realism is held constant in what Kania calls 'fundamental metaphysics'" (Dodd 2013: 1060).

In the third place, Dodd (2013: 1060–1061) stresses that there are cases in which folk beliefs about the nature of certain artefacts are radically wrong – this is e.g. true of the belief, held among the Ojibwa people, that dreamcatchers can actually catch bad dreams – and that therefore we cannot exclude that also our folk beliefs about the nature of artefacts such as musical works might be wrong.

Basing on the three criticisms just described, Dodd concludes that Kania fails to establish the determination thesis and therefore does not provide good reasons in support of local descriptivism in the ontology of music. Furthermore, Dodd (2013: 1064) shows that, if Kania were to appeal to Thomasson's arguments for the determination thesis, he would necessarily have to embrace Thomasson's global descriptivism, because (1) Thomasson stresses that her theory of reference-fixing must be applied to *all* kind terms (2007b: 64, 2007a: 190) and (2) according to Thomasson "difference between natural kinds and artefactual kinds exclusively concerns the determination of the specific manifest or empirically discoverable properties that are relevant to kind membership; it does not concern the determination of the ontological category to which such kinds belong" (Dodd 2013: 1064). The latter is a view that does not allow one to hold – as Kania would like to – that when we determine the ontological nature of artefacts and when we determine the ontological nature of other kinds of objects we operate in different ways, and that therefore we can be descriptivists about the ontology of artefacts and non-descriptivists about the ontology of other kinds of objects.

Letts' main criticism against Kania's local descriptivism is inspired by John Searle's remark that we should consider the possibility that the properties of certain artifacts, qua culture-dependent entities, be dependent upon the contents of our thoughts about them and that, at the same time, the contents of our thoughts misrepresent how things really stand (see Searle 2007: 17; Letts 2015: 218). For instance, it is possible that the dollar has certain value properties in virtue of our belief that there really is the right amount of gold underwriting it and that, however, this belief misrepresents how things really stand. Similarly, it would be possible for us to believe that musical works are created entities, that come into existence at a certain time, and that, however, this belief misrepresented the fact that musical works actually are eternal entities. Lett points out that Kania's view about the character of artifacts like works of music being determined by our practices would hold only if it could guarantee that the character of such artifacts never depends upon mistaken beliefs about them but, since this is not the case, we have a reason to reject the view.

To recap, Dodd's and Letts' criticisms show that Kania's local descriptivism (1) relies on a thesis – the determination thesis – that Kania fails to establish and (2) is undermined by the fact that it is incapable of excluding that our beliefs about works of art might be mistaken. Basing on Dodd's and Letts' arguments, then, it is safe to conclude that Kania's local descriptivism, as it stands, should be rejected.

16.2.3 Julian Dodd's Folk-Theoretic Modesty

Defending the view that works of music are eternal norm-types, Dodd (2007) champions a revisionist, anti-descriptivist ontology of art, based on two meta-ontological views: "meta-ontological realism" and "folk-theoretic modesty" (Dodd 2013: 1048–1049). Metaontological realism is the view that "the correct answers to first-order art-ontological questions – questions concerning the respective ontological categories the various artwork kinds belong to, their identity conditions, their persistence conditions, and so on – are objective [. . .] [in the sense that] their correctness is in no way determined by what we say or think about these questions" (Dodd 2013: 1048–1049). Folk-theoretic modesty is "the view that our common sense art ontological theories might substantially be mistaken" (Dodd 2013: 1048) and it follows from embracing meta-ontological realism: "For if the ontological nature of an artwork kind, *Fs*, is a matter of objective fact in the sense believed by the metaontological realist, then there is no guarantee that our common sense ontological conception of Fs is correct, even in outline" (Dodd 2013: 1049).

According to Dodd, then, in the ontology of music, like in standard metaphysics, we should maintain, in a Quinean fashion, that the correct answers to first-order art ontological questions should come from the objective domains of science and philosophical inquiry, rather than from the domain of our art-related practices: in other words, our ontology should stick to the semantic commitments of our best theories (Dodd 2013: 1051–1053). This standpoint, Dodd argues, conflicts with that occupied by all those theorists that he considers *descriptivists*: not only Thomasson and Kania, but also David Davies (2004, 2009, see next section), who allows for grounding the ontology of art on a rational reconstruction of artistic practices that "requires rational reflection from within our critical and appreciative practices, rather than the measuring of our critical judgments against an ontological theory"(Davies 2004: 22, see Dodd 2013: 1050–1051). Davies, however, has argued *contra* Dodd that "according a grounding role to artistic practice in the ontology of art does not conflict with the demands of meta-ontological realism and allows for both practices and beliefs about those practices to be revised" (Davies 2017: 124). I shall describe Davies' point in more detail in the next section.

16.2.4 David Davies' Rational Accountability View

Davies claims that the ontology of art should be "rationally accountable to our practices" and that, nevertheless, this "allows us to be revisionary" (Davies 2017: 127) and reject what Dodd (2013) has labelled the *determination thesis* (see § 16.2.2), i.e. the view that facts about works of art are not objective, but depend on the ontological beliefs implicit in our art-related practices. Davies' key point, which emerges from reflections on how we fix the reference of theoretical terms in science and how we should proceed in the analogous endeavour when concerned with art-ontological terms is that if

(a) we – just like Dodd – subscribe to folk-theoretic modesty and hold that if our folk-theories about works of art turn out to be false this nevertheless doesn't undermine our ability to talk about works of art (Davies 2017: 123);
(b) we – just like Dodd – argue that to decide which meta-ontology to adopt we should consider the virtues of the various theories on offer, such as "coherence with folk belief, but also norms such as *explanatory power*, simplicity, and integration with the findings of other domains" (Dodd 2013: 1051, my italics);
(c) and we, as a consequence of (a), "take terms that refer to particular artwork kinds as terms whose reference is fixed not 'descriptively' by our 'folk' beliefs or our 'folk practices' but by the capacity of their referents to play a particular explanatory role" (Davies 2017: 122);

then we must conclude that the explanatory role of the referents, in the context of ontological research on artworks, must be cashed out in terms of such referents' ability to explain our artistic practices. This, Davies stresses, is

> simply a recognition of the need to particularize any ontological inquiry to the things about which we are inquiring and a further recognition that what particularizes our ontological inquiry into the nature of artwork kinds is the explanatory roles that such kinds are intended to serve. The interests [in artworks as things that enter into certain human practices], in other words, determine *what it is whose ontological status is at issue*, not *what that ontological status is*. [. . .] a musical work, for example, must be something that stands in the right kind of causal-historical relation to our use of the term 'musical work' in our reflective musical practice and explains the salient features of that practice. [Davies 2017: 122–123]

Davies, then, does not claim that facts about musical works are not objective, but depend on the ontological beliefs implicit in our music-related practices (i.e. he does not subscribe to the *determination thesis*), and this allows him to reply, successfully, to the criticisms Dodd (2013: 1051–1058) addresses against previous descriptions of his meta-ontological view (Davies 2004; Davies 2009) – a view mistakenly labeled by Dodd as a form of *local descriptivism*.

What's the difference between simply adopting Dodd's *folk-theoretic modesty* and subscribing to Davies' *rational accountability view*, then? The difference lies in the results we get when it comes to our art-ontological views, as Davies explains (Davies 2017: 123–124). While Dodd favors a simpler ontology of artworks, according to which those works of art that are abstract objects are all eternal non-created abstracts, Davies is open to a pluralistic ontology of artworks, according to which those works of arts that Dodd understands as eternal abstract objects might actually belong to a variety of ontological kinds: on Davies' account, the ontology of classical musical works is likely to differ from the ontology of e.g. Balinese music. Davies thinks that his ontological outlook is superior to Dodd's because, if we accept the view that our art practices should rationally constrain our art ontology, then we should conclude that "the diversity of artistic practices is an argument against attempts to 'sanitize' ontology of art [like Dodd's one]" (Davies 2017: 124).

Davies, it seems to me, succeeds in showing that *folk-theoretic modesty* and the *rational accountability view* are compatible and therefore provides a good competitor to Dodd's meta-ontological view. Given the flaws of Thomasson's and Kania's views, Dodd's and Davies' proposals on the meta-ontology of art emerge as the most interesting currently on offer.

16.3 Conclusion: The Advantages of Hermeneutic Fictionalism

In Sect. 16.1, I have defended a fresh proposal about the meta-ontology of artworks such as works of music. In a nutshell, I have argued that we can always let our folk statements about musical works guide our research in the ontology of music: since it can be argued that the existence of musical works is not part of the asserted content of our folk talk about musical works, then we should not commit ourselves to the inclusion of musical works into our ontology. Rather, we should claim that when we talk about musical works we are talking fictionally.

My hermeneutic fictionalist approach to the ontology of musical works, inspired by a descriptivist approach to the semantics of folk musical works talk, shows a sense in which our ontology of artworks can remain faithful to our art practices and related art talk – a sense which has not been grasped by other scholars: we can be faithful to our art talk while researching the ontology of artworks if we take into account what is perceived as irrelevant by speakers who talk about e.g. works of music, i.e. the issue whether such works exist or not. In particular, my view has a significant advantage over Dodd's and Davies' ones: hermeneutic fictionalism does not require us to subvert, if necessary, common sense opinion about works of art, and it does not require that we embrace an error theory about e.g. our folk music talk. What it proposes is that we have always been talking metaphorically when talking about musical works and that there is no need to distance ourselves from our metaphorical talk, which is, instead, expressively useful.

Finally, the view I have sketched out not only embraces meta-ontological realism[12] (like Dodd's and Davies' views), while respecting the practice of folk

[12]Recall that metaontological realism is the view that "the correct answers to first-order art-ontological questions – questions concerning the respective ontological categories the various artwork kinds belong to, their identity conditions, their persistence conditions, and so on – are

talk about works of art (in a way that escapes both Dodd and Davies), but it also shows a way out of the impasse reached by the debate between Dodd and Davies. Since we don't have to commit ourselves to the existence of works of music, we are not required to choose whether to privilege ontological simplicity at the expense of explanatory power, opting for Dodd's folk-theoretic modesty and conceiving of e.g. works of music as eternal abstract objects, or to privilege explanatory power at the expense of a simpler ontology, opting for Davies' rational accountability view, and conceiving of e.g. works of music as created abstract artifacts.[13]

References

Boghossian, P. (1994). Inferential role semantics and the analytic/ synthetic distinction. *Philosophical Studies, 73*, 109–122.

Burgess, J. (1983). Why I am not a nominalist. *Notre Dame Journal of Formal Logic, 24*, 93–105.

Burgess, J. (2008). Numbers and ideas. In *Mathematics, models and modality: Selected philosophical essays*. Cambridge: Cambridge University Press.

Burgess, J., & Rosen, G. (1997). *A subject with no object: Strategies for nominalistic interpretations of mathematics*. Oxford: Clarendon Press.

Caplan, B., & Matheson, C. (2006). Defending musical Perdurantism. *BJA, 46*(1), 59–69.

Davies, D. (2004). *Art as performance*. Oxford: Blackwell.

Davies, D. (2009). The primacy of practice in the ontology of art. *JAAC, 67*(2), 159–171.

Davies, D. (2017). Descriptivism and its discontents. *JAAC, 75*(2), 117–129.

Devitt, M. (1981). *Designation*. New York: Columbia University Press.

Dodd, J. (2007). *Works of music*. Oxford: Oxford University Press.

Dodd, J. (2012). Defending the discovery model in the ontology of art: A reply to Amie Thomasson. *BJA, 52*(1), 75–95.

Dodd, J. (2013). Adventures in the metaontology of art: Local descriptivism, artefacts and dreamcatchers. *Philosophical Studies, 165*, 1047–1068.

Evans, G. (1973). The causal theory of names. *Proceedings of the Aristotelian Society, Supplementary Volume, 47*, 187–208.

Frege, G. (2007, ed. or 1884). *The foundations of arithmetic* (D. Jacquette, Trans.). New York: Longman.

Irvin, S. (2012). Installation art and performance: A shared ontology. In C. M. Uidhir (Ed.), *Art & abstract objects*. Oxford: OUP.

Kania, A. (2008a). Piece for the end of time: In defense of musical ontology. *BJA, 48*(1), 65–79.

Kania, A. (2008b). The methodology of musical ontology: Descriptivism and its implications. *BJA, 48*(4), 426–444.

Kania, A. (2012). Platonism vs. nominalism in contemporary musical ontology. In C. MagUidhir (Ed.), *Art and abstract objects*. Oxford: OUP.

objective [. . .] [in the sense that] their correctness is in no way determined by what we say or think about these questions" (Dodd 2013: 1048–1049).

[13] I am deeply thankful to David Davies, Jerrold Levinson and Matteo Plebani for their comments on previous versions of this paper and to audiences at the "Abstract Objects" Workshop at the University of Santiago de Compostela (2016), the Conference of the Italian Society for Analytic Philosophy (Pistoia 2016), the Conference of the European Society for Analytic Philosophy (Munich 2017), and the Annual Conference of the American Society for Aesthetics (Toronto 2018) for their valuable feedback on previous versions of this paper.

Letts, P. (2015). Against Kania's fictionalism about musical works. *BJA, 55*(2), 209–224.
Levinson, J. (2012). Indication, abstraction, and individuation. In C. MagUidhir (Ed.), *Art and abstract objects*. Oxford: OUP.
Liggins, D. (2010). Epistemological objections to platonism. *Philosophy Compass, 5*(1), 67–77.
Liggins, D. (2011). *"Fictionalism", Oxford bibliographies online*. New York: Oxford University Press.
Liggins, D. (2014). Abstract expressionism and the communication problem. *British Journal for the Philosophy of Science., 65*, 599–620.
Putnam, H. (1975). *The meaning of 'meaning'* (Minnesota studies in the philosophy of science) (Vol. 7, pp. 131–193). Minneapolis: University of Minnesota Press.
Quine, W. V. (1948). On what there is. *The Review of Metaphysics, 2*(5), 21–38.
Reimer, M., & Michaelson, E. (2017). Reference. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2017 Edition). https://plato.stanford.edu/archives/spr2017/entries/reference/
Ridley, A. (2003). Against musical ontology. *The Journal of Philosophy, 100*, 203–220.
Searle, J. (2007). Social ontology and the philosophy of society. In E. Margolis & S. Laurence (Eds.), *Creations of the mind: Theories of artifacts and their representation*. Oxford: OUP.
Stanley, J. (2001). Hermeneutic fictionalism. In French & Wettstein (Eds.), *Midwest studies in philosophy volume XXV: Figurative language* (pp. 36–71). Oxford: Blackwell.
Thomasson, A. (2004). The ontology of art. In P. Kivy (Ed.), *The Blackwell guide to aesthetics* (pp. 78–92). Oxford: Blackwell.
Thomasson, A. (2005). The ontology of art and knowledge in aesthetics. *JAAC, 63*(3), 221–229.
Thomasson, A. (2006). Debates about the ontology of art: What are we doing here? *Philosophy Compass, 1*, 245–255.
Thomasson, A. (2007a). *Ordinary objects*. Oxford: OUP.
Thomasson, A. (2007b). Artefacts and human concepts. In E. Margolis & S. Laurence (Eds.), *Creations of the mind: Theories of artefacts and their representations*. Oxford: OUP.
Walton, K. (1990). *Mimesis as make-believe: On the foundations of the representational arts*. Cambridge: Harvard University Press.
Walton, K. (1993). Metaphor and prop oriented make-believe. *European Journal of Philosophy, 1*(1), 39–57.
Yablo, S. (1998). Does ontology rest on a mistake? *Proceedings of the Aristotelian Society, 72*, 229–226.
Yablo, S. (2001). Go figure: A path through fictionalism. *Midwestern Studies in Philosophy, 25*, 72–102.
Yablo, S. (2002). Abstract objects: A case study. *Philosophical Issues, 12*, 220–240.
Yablo, S. (2009). Must existence-questions have answers? In D. J. Chalmers, D. Manley, & R. Wasserman (Eds.), *Metametaphysics: New essays on the foundations of ontology* (pp. 507–525). Oxford: OUP.
Yablo, S. (2010). *Things*. Oxford: OUP.
Yablo, S. (2014). Carnap's paradox and easy ontology. *The Journal of Philosophy, 111*(9/10), 470–501.

Chapter 17
A Realist-Friendly Argument for Moral Fictionalism: Perhaps You'd Better Not Believe It

Christopher Jay

Abstract We don't have to choose between fictionalism and realism. There would, perhaps, be something good about our acceptance of moral claims – our moral commitments – being nondoxastic, i.e., amounting to some form of acceptance other than belief, even though the moral claims accepted are apt to be believed (i.e. their semantic content is fully representational), and even if they are *true*. I present an argument to this conclusion which does not rely upon any non-realist assumptions, and which is in fact strengthened by making some realist assumptions. As well as being an independently interesting argument, then, it shows that realists might have reasons to be fictionalists too.

Keywords Fictionalism · Moral realism · Moral beliefs · Moral commitment · Epistemic Akrasia

17.1 Introduction

Whatever moral facts or truths are or would be, they presumably don't depend upon the existence of distinctively moral *objects* in the way that mathematical facts might be thought to depend upon the existence of objects such as numbers or sets. So, the issues which are in play when it comes to moral fictionalism are slightly different from the issues in play when it comes to mathematical fictionalism, or fictionalism about possible worlds or composite objects etc., and this chapter therefore does not deal with abstract objects as such. It does, however, touch on moral realism and its relation to fictionalism. Most discussions of fictionalism start from the rejection of realism about some domain and progress to a discussion of the attitudes appropriate in light of that rejection, and most discussions of moral fictionalism start accordingly from the rejection of moral realism. But this chapter presents an argument for moral

C. Jay (✉)
Department of Philosophy, University of York, York, UK
e-mail: christopher.jay@york.ac.uk

J. L. Falguera, C. Martínez-Vidal (eds.), *Abstract Objects*, Synthese Library 422,
https://doi.org/10.1007/978-3-030-38242-1_17

fictionalism which is available to a moral realist. A moral fictionalist need not be against moral realism, and a moral realist need not reject moral fictionalism.

This is an essay on the ethics of *nondoxastic commitment*, in the tradition of discussions of the ethics of belief such as the one famously offered by W. D. Clifford.[1] Our beliefs are of moral concern at least in part because they are action guiding. And that feature of beliefs is shared with other sorts of commitments. Just as our moral beliefs, and our beliefs about the non-moral facts, inform our deliberation and decision with respect to action, so our *nondoxastic* moral, political and religious commitments play much the same roles, or at least they do if we have any.[2] So an ethics of commitment which only has something to say about beliefs is incomplete.

Nondoxastic commitments are psychological states which involve *accepting* some proposition but not *believing* it. Commonly this possibility is illustrated by pointing out that a great many people accept that Sherlock Holmes was a detective without, strictly speaking, believing it: they are willing to say that Sherlock Holmes was a detective, for example, but they know very well that he was not, since Sherlock Holmes did not exist, and (as they well know) nothing which did not exist could have actually been a detective.[3] In the philosophy of mathematics, philosophy of science, philosophy of religion and metaethics, arguments have been advanced which appeal to the idea that we can accept some proposition without believing it for various pragmatic reasons, and it is sometimes argued that whether or not this is what we do as things stand, it is what we *ought* to do. Fictionalism is the view that our commitments in some domain are or ought to be (or, as I'll argue below, might well be) nondoxastic, although – and this is what distinguishes it from canonical non-cognitivism (see Kalderon (2005a, b) – fictionalists agree with traditional realist cognitivists that the *semantic content* of what is accepted is fully representational, and should be taken literally: fictionalism (as I understand it here) does not reinterpret its target claims; it adopts a perfectly 'standard' non-expressivist

[1] See Clifford (1876–7). Clifford would not, however, agree with much of what I go on to say in this chapter, which might usefully be read alongside Adams (1995).

[2] This claim is not uncontroversial, but I think it is true and will assume it here. Some philosophers doubt that the *rationality* of action can be explained (even in part, apparently) by appeal to nondoxastic commitments, but suffice it for now to point out that plenty of appealing theories of action and rationality allow a role for 'acceptance' which is nondoxastic: see, e.g., Bratman (1992) (though the sort of nondoxastic acceptance I have in mind is, as will emerge, rather different from the sort Bratman has in mind), or Velleman (2000). See also Joyce (2005).

[3] You might think that in fact what is going on is that they are just believing a proposition about what some stories say, rather than nondoxastically accepting the (different) proposition that Sherlock Holmes was a detective. But see Joyce (2005) for good reasons not to think that. As it happens, I think the fiction analogy is unhelpful in some ways – see Jay (2012: Chap. 1) for discussion – but I employ it here for convenience since my purpose is not to provide a full-blooded discussion of all aspects of fictionalist positions (see Jay (2012: Part 1) for more).

and literal account of their meaning, but insists on the fact that belief is not the only propositional attitude which can be adopted in accepting claims with that sort of meaning.[4]

Fictionalists who argue that we ought to have nondoxastic moral commitments argue that there is something *good* about our having nondoxastic commitments. But all sorts of things are good in some respect, without it following that we ought to do or have them, and having identified something good about nondoxastic commitments there remains work to be done. The fictionalist who thinks that our commitments *ought* to be nondoxastic must establish (or rely upon the independent plausibility of) two distinct claims: firstly, that having *some* commitments with respect to the relevant domain is better than having *none*; and, secondly, that it is better to have *nondoxastic* commitments with respect to that domain than to have (just) *beliefs*.[5]

A good example of this strategy in metaethics is Richard Joyce's argument for moral fictionalism. (See Joyce 2001: Chaps. 7 and 8; 2005.) Joyce argues that our moral commitments are extremely valuable bolsters to our resolve, allowing us to do things which we are not immediately inclined to do, and to resist weakness of will. To have some moral commitments, then, is better than to have none. But Joyce also thinks that the moral claims we accept in having moral commitments – which for most of us as things stand are moral beliefs – are systematically untrue, for Joyce is an error theorist. So, given that it is bad to have untrue beliefs, it is better to have *nondoxastic* moral commitments than moral beliefs, for there is nothing necessarily bad about having untrue nondoxastic commitments. The upshot, then, is that there is good reason for us to have nondoxastic moral commitments as opposed to either moral beliefs or no moral commitments at all.[6]

[4]This sense of 'fictionalism' and its cognates (which Kalderon describes as 'non-cognitivism without non-factualism') is not the only one in circulation. Rosen's (1990) fictionalism, for example, is rather different: as I am thinking of it, fictionalism does not involve a view of its target claims as involving any sort of tacit *according to...* operator. The issue here relates to that discussed in the previous footnote. See Joyce (2005), or my (2012: pp. 48–51). For a similar understanding of fictionalism (though a different understanding of 'realism' from the one relied upon here), see e.g. Kroon (2011: p. 787).

[5]Of course, it does not immediately follow even from this that we *ought* to have nondoxastic commitments rather than beliefs, or that we ought to have nondoxastic commitments at all: it might be that, as things stand, there are more or less (or completely) insuperable practical obstacles, or unacceptable practical costs, to shifting from belief to nondoxastic acceptance (if beliefs about the relevant domain are what we currently have). This is why (as I argue in my (2014: §2)) thinking of fictionalism as either *hermeneutic* (claiming that our commitments actually are nondoxastic) or *revolutionary* (claiming that out commitments are not but ought to be nondoxastic) is too simplistic. The argument I give in this paper will be for what I call *evaluative* fictionalism, claiming that nondoxastic commitment would be *good*. This does not entail the *normative* conclusion of revolutionary fictionalism, and it is silent with respect to the *descriptive* claim of hermeneutic fictionalism.

[6]The structure of this argument is similar to Field's (1980, 1989) argument for mathematical fictionalism, and to an argument which, I argue, is recoverable from Kant's work on the theological postulates (see my (2014)).

In what follows, I will develop an argument with the same structure, to the conclusion that there is something particularly good about nondoxastic moral commitment. I do not go so far as claiming that we ought to have nondoxastic moral commitments as opposed to moral beliefs, because I will not consider the *dis*advantages. The point of presenting my argument is not to settle that issue, but to suggest that there is something to be said for nondoxastic moral commitment the force of which has not been acknowledged so far in the debate but which deserves to be.

The good in the nondoxastic which I will be highlighting is interesting because it is a good which might persuade a moral *realist* to embrace fictionalism. As I said above, it is generally assumed that fictionalism is a position which might appeal only to non-realists.[7] In my intended sense, a moral realist is someone who thinks not only that our moral claims are truth-*apt* but also that a significant class of moral propositions are in fact *true*.[8] Non-realists, then, fall into two classes: those who *deny* that a significant class of moral propositions are true (because they take moral claims either to fail of truth-aptness or to be false); and those who think that we are merely in no position to assert that they are true (though they also think we are in no position to assert that they are false or (less likely) that they fail of truth-aptness). Fictionalists in various domains are split between the ‘atheists’ (e.g. Field (1980, 1989) and Leng (2010) with respect to mathematical objects) and the ‘agnostics’ (e.g. van Fraassen (1980) with respect to scientific unobservables), and in each case it is easy to see why, being unwilling to embrace realism, they are drawn to a view which approves of nondoxastic commitments which promise to retain what is good about commitment with respect to that domain whilst avoiding untrue – or at least unwarranted – beliefs.

It is often assumed that even if fictionalism is *compatible* with realism, the realist has no *reason* to embrace fictionalism, since if beliefs are available and respectable in virtue of their being true, they will presumably be able to do whatever nondoxastic commitments can do, and probably do it better.[9] But if the argument I present below is along the right lines, this assumption is unjustified, for there might very well be some good which our nondoxastic moral commitments are in a position to secure and our moral beliefs are not (at least to the same extent).

[7]I present realist arguments for *religious* fictionalism in my (2014) and (2016).

[8]The ‘significant class’ qualification is intended to rule out counting as a moral realist in virtue of thinking that just the *negative* moral claims (e.g. *gratitude is not wrong*) are true, and true *vacuously* (because *nothing* is wrong).

[9]For discussion of the *formal compatibility* of realism with fictionalism, see my (2012: Part 1, esp. chs1, 2). That realism is at least formally compatible with fictionalism has been acknowledge more or less in passing by Brock (2002), Nolan (2005), Yablo (2002: esp. fn1 and sec.12; 2005: fn), and, perhaps, Kalderon (2005a) who says ‘[m]oral fictionalism is consistent with the existence of moral facts’ (p. 179).

17.2 Subjective Warrant and Action

Epistemic warrant, as I will understand it, is that which speaks in favour of believing that *p*. (That which *epistemically* speaks in favour of believing that *p*, that is: not that which, for example, pragmatically speaks in favour of doing so.) It does not follow straight from the definition of epistemic warrant that an *un*warranted belief is *defective*: it might be that there is not necessarily anything defective about a belief of which nothing speaks in favour (so long as nothing speaks *against* it).[10]

Whether or not there is objective epistemic warrant for a belief is a matter of whether or not what is supposed to speak in favour of that belief is in fact the case (whether a believer has some evidence, for example). But a person can *take there to be* objective epistemic warrant when there is in fact none, as when, for example, I think that there is evidence that *p* but the facts I have in mind do not stand in the evidential relation to *p*, or when the putative facts I have in mind which would stand in that relation do not obtain.

A person's *taking there to be* objective warrant for their belief is their having *subjective* epistemic warrant: objective epistemic warrant is what epistemically grounds a belief (or, perhaps better, a person's believing), and a person's subjective epistemic warrant for that belief is their thinking that they have those grounds.

Note that degree of subjective warrant is a matter of one's confidence in the *grounds* of one's belief, not a matter of the *credence* associated with that belief. As I am thinking of these issues, a person can *fully* believe that *p* despite their lacking confidence in the well-groundedness of that belief. It is, of course, a further matter whether it is ever *rational* for someone to be in this state. The 'can' in the claim I have made is the 'can' of psychological possibility, not the 'can' of rational permissibility, and I take this psychological claim to be plausible according to at least some conceptions of the nature of belief.

Having isolated a notion of subjective epistemic warrant for beliefs, I want to now broaden the notion of subjective warrant to cover commitments more generally. Warrant – subjective and objective – for beliefs is *epistemic* just because of the way beliefs are truth or evidence normed: the grounds for belief, which speak in favour of believing, are going to be connected to truth or to truth-related notions such as evidence. The warrant for nondoxastic commitments will not typically be epistemic, though, for nondoxastic commitments are not typically truth or evidence normed.

What counts as objective warrant for some nondoxastic commitment depends upon what in fact amounts to appropriate grounds for that type of commitment. I think nondoxastic commitments come in many varieties, and they are appropriately grounded in different ways. For now, let's call the appropriate grounds for some commitment – the grounds which confer objective warrant – *C*. Whether some commitment is objectively warranted is just a matter of whether *C* obtains. Whether a person, *A*, has subjective warrant for some commitment will then be just a matter of whether *A* believes with some confidence that *C* obtains.

---

[10] See Harman (1986, 1999).

With this characterisation of subjective warrant in hand, we are ready to make the point I will rely on in my fictionalist argument. It is a point about the relation between subjective warrant and action.

Even though our beliefs are sensitive to the available evidence in the ideal case, we often have dogmatic beliefs. And as well as dogmatic belief, which refuses to budge despite the evidence, there is what I'll call *hyper-enkratic* belief. Hyper-enkratic beliefs are those which are not only dogmatically maintained but which a person is also willing to *act on* (to form intentions based on) despite their lack of evidence (objective warrant). Someone who maintains beliefs about the rights and wrongs of something despite good reasons to abandon or at least suspend those beliefs further to reconsidering the available evidence is dogmatic. Someone who goes further and acts on – or intends to act on – their dogmatic beliefs is hyper-enkratic.

I do not want to underplay the prevalence of hyper-enkratic action. But I also want to suggest that there is a converse condition, the condition of *sub-enkratic* belief, which consists in a person not acting on fully held beliefs and in their failing to act on them for a very particular reason: the sub-enkratic person fails to form an intention to act on their belief (despite their fully believing) *because they lack subjective warrant* for that belief.[11]

Consider, for example, the following case. Alf knows that Betty is having an affair with George, behind her husband Wilfred's back. Alf and Betty are friends, whereas Alf and Wilfred have got along well enough without ever really becoming close. Alf (fully) believes that what he ought to do is tell Wilfred about Betty's shenanigans, because he (fully) believes that a chap has a right to know that he is cuckolded and that, regardless of personal loyalties, anyone in possession of the relevant information is under an obligation to give the cuckold his due, namely the truth. But Alf knows that he believes this because he was brought up in a time when sexual and marital ethics were very different from what they are now, and he thinks that if he were a young man *now*, he would doubtless have a different view and would probably be inclined to put his loyalty to his friend, Betty, ahead of what he currently thinks of as his duty to his mere acquaintance Wilfred. Whilst this knowledge does not make him believe what he believes any less, it does make him seriously wonder whether his grounds for those beliefs are as good as they might be.

In some cases, Alf is willing to put his doubts about the grounds of his moral beliefs aside and act on them nonetheless, because for all their faults they are at least his genuine moral convictions. So, he is willing to tell children off in the street

[11] I reserve the term 'sub-enkratic' and its cognates for this particular phenomenon in this paper. But it is probably as well to call any beliefs which do not issue in the intentions they typically would, for whatever reason (including reasons which are nothing to do with lack of subjective warrant) 'sub-enkratic' in other contexts.

for swearing and jumping bus queues, even though he strongly suspects that his moral attitudes towards these things are more a product of his stiff pseudo-Victorian upbringing than anything else (and even though those very attitudes embarrass him slightly), because good old fashioned moral attitudes like those are part of what make him *him*, and because the stakes are low. On the other hand, when what is at stake is the happiness and wellbeing of both his friend, Betty, and two other people it would be quite understandable if Alf's resolve to act on his moral beliefs were to falter. Without ever feeling less confident that telling Wilfred of Betty's antics is the right thing to do, Alf might very well decide, no doubt with great anguish, that that belief is just too ill-grounded to be the basis for his intervention. Were Alf's confidence in the grounds of his belief to be restored, or were he to discover new and impressive grounds for believing what he already believes, then we must suppose that he would decide that Wilfred must be told, because Alf takes morality very seriously, and it was never the strength, but only the grounds, of his moral beliefs which were in doubt. He would, of course, still regret the harm done to Betty and the others by telling Wilfred all, but it is not *this* concern which stays his hand as things stand. It is rather the simple fact that he lacks subjective warrant for his moral beliefs.

If, as I think, the story of Alf is at all psychologically plausible, it suggests that in our high-stakes deliberations about what to do the *strength* and *content* of our moral beliefs are not the only salient considerations: there is also our *confidence in their grounds*.

It might be argued that it is *irrational* to take one's belief that *p* to be ill-grounded – or to suspect it of being so – but to fail to revise or abandon it. But, firstly, it is not obviously irrational to maintain a belief for which one has lost (or never had) subjective warrant: perhaps it is irrational to fully uphold it by intending to *act* on it; but it might not be irrational to believe that *p* on (what one takes to be) *no* good grounds, so long as there are no better grounds for believing that *not-p*.[12] And, secondly, even if it is the case that it would be irrational to maintain a belief for which one has lost (or never had) subjective warrant, it most certainly does not follow that maintaining such beliefs is never something we do: we are,

[12]There is a substantial literature concerned with so-called 'higher order evidence', and the epistemology of self-doubt and the rational response to reasons for self-doubt more generally (see Roush 2017). It might be supposed that the plausibility of my argument depends upon what the right thing to think is about the rational permissibility of so-called 'level splitting', in which a person continues to believe *p* when they (knowingly or believingly) have higher order evidence or other reasons for thinking that their evidence for or reasons for believing *p* are less reliable than their belief in *p* requires. The literature on higher order evidence and related issues in the epistemology of self-doubt is far from unanimous in its judgement here. That at least some cases of such level splitting would be rationally permissible *might* be allowed by several prominent participants in the debate: see Roush (2017: esp. §3), citing e.g. Williamson (2011), Weatherson (2008), Coates (2012) and Wedgwood (2011). But see the next point in the main text, which explains why it doesn't really matter who is right about the *rationality* issue for the purposes of my argument.

unfortunately, all too prone to fall short of the demands of rationality, and if we *do* continue to fully believe what we lack subjective warrant for (albeit irrationally), the issue I am interested in – about the relation between subjective warrant and action – arises.[13]

I do not claim that lack of subjective warrant for a moral belief *necessarily* or *always* undermines that belief's role in forming intentions even in high stakes situations, for I am aware that hyper-enkratic action is common. We can all think of cases where we suspect that someone has acted on a belief for which they lack even subjective warrant, perhaps because saving face requires following through with a plan despite losing confidence in its merits, or perhaps because of a sincere belief that some action is better than none, and this is the only thing they can think of doing however worried they are about their reasons. Whether it turns out that people ever fail to form intentions based on full beliefs because of lack of subjective warrant is, surely, an empirical question, and so we must not overreach ourselves. But I hope that the case I have just presented makes it plausible that at least some moral beliefs *are* sub-enkratic. There is no reason to think that human psychology is so neatly ordered that we only make one type of mistake: sometimes we are too quick to act; sometimes too reticent.

Before moving on to the next plank of my argument, I want to emphasise that nothing I have said about the relationship between subjective warrant and action rests on whether a person's reasons for doubting the grounds of their beliefs are good or not. So, when faced with Alf's story, you might think that Alf's doubts about the grounds of his beliefs are rather silly, so you might be saddened by Alf's loss of resolve. But you will have no reason thereby to doubt that what I have claimed about the relationship between subjective warrant and action is true. I will return to the fact that loss of subjective warrant can itself be (objectively) unwarranted in what follows.

[13] Note that the literature on higher order evidence and related issues in the epistemology of self-doubt is more or less entirely concerned with the *rationality* of 'level splitting' or the rational way to respond to self-doubt; it is more or less silent, as far as I can tell, on the question of whether it is *psychologically possible* to be self-doubting in the way I have described. Harman (1986: Chap. 4) argues that it is 'incoherent to believe both *P* and also that all one's reasons for believing *P* relied crucially on false assumptions'. There is a sense in which one's beliefs, in such a case, obviously do not cohere: there is clearly a tension. But if 'incoherent' is supposed to mean anything stronger than that, and in particular if the suggestion is that it is incoherent to suggest that we are sometimes in such a state, we should, I think, be sceptical of this claim. See Roush (2017: esp. §1), however, for discussion of some views according to which strict consistency and coherence requirements might condemn the kind of self-doubt we are concerned with here. Once again, though, these views are concerned with the *rationality*, not the possibility, of being or remaining in such a self-doubting situation; so even if correct, they do not undermine the crucial point relied upon here.

17.3 Scepticism and Other Threats to Subjective Warrant

We have seen, in the story of Alf, one way of losing subjective warrant for one's strongly held moral beliefs. In Alf's case, it was the worry – Alf's own worry, which may or may not be reasonable – that, though his moral beliefs are unshakable and strong, his reasons for believing what he does are no better than the reasons others have for believing very different and incompatible things. Alf's attitude to the grounds of his moral beliefs is not informed by clever arguments or high-flown theories. He simply notices that others are very different from him in what they believe about morality, and he can think of nothing to reassure himself that they have not got it at least as right as he has, though he is far too committed to his existing beliefs to abandon them. Alf is simply *modest*.[14]

But you do not have to be modest in the way Alf is to lose subjective warrant for your moral beliefs. You might be convinced by a clever *argument* instead that you lack objective warrant for your moral beliefs (whilst retaining those beliefs nonetheless). The arguments I have in mind are those which purport to show (in decreasing order of strength) that there is nothing in the way of moral truth (of an interestingly robust sort) for there to be knowledge *of*, and/or that we have no such knowledge, and/or that we at least have no objective warrant for our moral beliefs, and/or that at least we are in no position to take ourselves to have such objective warrant. As is familiar, there are all sorts of arguments which purport to establish such conclusions.[15] And to become convinced by such arguments is, if one

[14]For a philosophical presentation of this sort of modesty worry about 'nurtured beliefs', see Cohen (2000: Ch. 1).

[15]It might be worth noting that some sceptical arguments (of what I'll call the *anthropological* and *genealogical* types) amount to philosophical responses to the sort of sensitivity to contingency which is at the root of the modesty view I have also mentioned, thus bridging the gap between what I have described as two sources for loss of subjective epistemic warrant. Anthropological arguments (such as the 'argument from relativity' which Mackie (1977: Ch1, §8) suggests but does not ultimately rely on) use the supposed fact that there is pervasive moral disagreement (between individuals or, more seriously, between cultures or times), or the use of radically different moral concepts (as in Williams [1975]), to argue that it is at least unlikely, if not unintelligible, that there is an objective truth in the vicinity, or that we are grasping it. Genealogical arguments do not rely on there in fact being any variation in moral judgement, but instead appeal to the supposedly non-truth-tracking nature of how we came to form moral judgements to argue that we are, at best, lucky if they are veridical and perhaps even that we *could not* form veridical moral judgements on the basis of the sorts of processes which have got us to where we are. A classic genealogical argument is in Nietzsche [1887], according to a popular reading of that text (offered by, for example, Sinhababu (2007)) at least. Recently, genealogical debunking arguments have tended to appeal to evolutionary psychology – see, e.g., Joyce (2006), Street (2006), and Kahane (2011) for discussion. As I said, I mention these sorts of arguments because they (like the argument discussed by Cohen, cited in the previous footnote) are philosophical reflections of the sort of ordinary modesty I have also appealed to; I do not mean to suggest that these are the most promising sorts of sceptical arguments, much less that they are successful. For a good collection of essays, which is it useful to compare with my argument here, exploring what the right philosophical response might be to genealogical arguments which show that morality is a mere 'ideology' see Harcourt (2000).

is rational, to take oneself to lack objective warrant for any moral beliefs which one has, or, in the terminology I have been employing, to lose subjective warrant for those beliefs.

Again, the argument I am developing does not require me to take a stand on whether these arguments are good. The point is just that however things stand with respect to the quality of such arguments, those who are convinced by them do not always give up – or even weaken – their first-order moral beliefs. Again, lack of subjective warrant does not *destroy* belief, even if it casts particular beliefs in a dubious rational light. Many people who come under the spell of sceptical arguments simply retain their beliefs, and adopt an error theory with respect to them, taking them to be lacking warrant even as they inform their deliberation and action, and this is as true with respect to beliefs about causation, the external world and aesthetic qualities which are condemned by sceptical arguments as it is with respect to beliefs about goodness, duty or virtue.

There are, then, various ways of (or reasons for) coming to lack subjective warrant for our moral beliefs. If the claim I defended in the previous section is true, then coming to lack subjective warrant for those beliefs might undermine our forming intentions based on them as we ordinarily would – if not in all contexts, then at least in some high stakes situations. It is an upshot of this that certain forms of modesty or sympathy towards sceptical arguments might lead us, in a distinctive way, to fail to act as we ought to.

17.4 The Problem with Belief, and the Fictionalist Fix

Certain forms of modesty and sympathy towards sceptical arguments might lead us to fail to act as we ought to in a *distinctive* way because it is not, as is so often the case, that the explanation for our failure is either (i) lack of insight or (ii) weakness of will. These are often sufficient explanations of our moral failings. But the person who has simply lost confidence in the *grounds* of their moral belief might have the *right* belief (so might not lack insight), and it might take considerable strength of will *not* to act on it (as when allowing something goes against everything one stands for).

The reason this is the basis for a fictionalist argument is that whilst lack of subjective warrant is undermining with respect to the action guiding role of any moral commitment, modesty and scepticism threaten the subjective warrant only of *beliefs*. It is subjective *epistemic* warrant that is undermined by reflecting upon the (supposed) fact that others with commitments which are inconsistent with your own have reasons for their commitments which are just as good as your reasons for yours, or by thinking that one's commitments fail to track the truth. If our moral commitments were such as to require no subjective *epistemic* warrant for their being fully upheld as the source of intentions, then modesty and scepticism would not affect them in the way they affect belief (for which, by definition, warrant is epistemic). And nondoxastic attitudes towards morality, in which moral claims are *accepted* but *not believed*, are exactly what moral fictionalists think our moral attitudes might consist in.

But isn't that just to say that our moral commitments might amount to wishful thinking? I think not, or at least not in any pejorative sense of 'wishful thinking'. The pejorative sense of 'wishful thinking' is the sense in which wishful thinking amounts to *believing* based on our desire to believe, or based on our desire for something to be true. Wishful thinking in this sense is different from nondoxastic moral commitment in that belief lacks warrant if what speaks in its favour is just that we would like to believe, or that we would like something to be true. That is part of what it is for belief to be governed by epistemic norms: *wanting* to believe *p* or *wanting p* to be true are not evidence that *p*, so there is something wrong with *believing* on the basis of wanting to believe or on the basis of wanting something to be true. But wanting to have some other sorts of commitments, and embracing those commitments for that reason, is not necessarily wrong at all. It is not even necessarily wrong if I know that those commitments involve accepting untrue or epistemically unwarranted claims. That is what it means to say that the norms of those commitments are non-epistemic, but rather pragmatic, aesthetic or whatever.

It often matters, morally, if we fail to do what we ought to do. So, on the assumption that a person who lacks *all* moral commitments is unlikely to act well (perhaps because acting well just *is* (perhaps *inter alia*) acting guided by moral considerations, or perhaps just because it is very unlikely that a person with no moral commitments would chance upon the right choice, in hard cases at least), it matters, morally, that we have some moral commitments rather than none. The claim that it matters that we have some moral commitments rather than none is, you will recall, one of the claims that the fictionalist needs to establish. And I take it to be plausible not just on the grounds which I have already mentioned (to do with the likelihood of *acting* well in the absence of moral commitments) but also perhaps because a person without moral commitments is deficient in a morally important way regardless of what their lack of moral commitments means for their actions. Plausibly, it matters whether a person recognises any moral constraints on their will, or any moral claims upon them, whether or not their failing to recognise such things would undermine their acting in the right ways. So, I take it there are various good reasons to think that if moral beliefs turn out to be problematic in some way, it is preferable to have some nondoxastic moral commitments rather than none, for at least nondoxastic commitments are commitments.

Some philosophers doubt that nondoxastic commitments could play a morally valuable action-guiding role. This might be because they assume that nondoxastic commitments are bound to be *flimsy*, which is to say that they are too liable to be given up when the going gets tough to be reliably action guiding. But it is implausible that nondoxastic commitments are flimsy just in virtue of being nondoxastic. Think, for example, of a person's nondoxastic commitment to Sherlock Holmes being a detective.[16] It seems very unlikely that they would be talked, cajoled or tricked into replacing that commitment with the commitment that Sherlock Holmes was *not* a detective, or into abandoning their commitment. Perhaps if it came to light that there is a Conan Doyle story featuring Sherlock Holmes in which it is made clear that this Holmes figure is merely masquerading as a detective, and is in fact a dreadful villain playing a confidence trick on the hapless Watson, they would nondoxastically accept that he is not a detective after all. But to the extent that the emergence of *new grounds* is liable to undermine what one accepts, belief is no less flimsy than nondoxastic forms of commitment: new evidence is liable to get someone to change their beliefs in just the same way. I do not mean to say that there is no possibility of nondoxastic moral commitment being flimsier than belief; but I do mean to put the burden of argument on those who assume that it is, for it is not a *general* truth about nondoxastic forms of commitment that they are flimsier than beliefs.

Another reason why some philosophers think of nondoxastic commitments as deficient in respect of their action guiding roles might be that they worry about their *motivational strength*. The motivational strength worry is that whilst a person might have very stable nondoxastic commitments, those commitments are not going to get a person to act in the way belief can. Perhaps the thought is that it is one thing to act because you believe that what you are doing is right, but quite another to act just because you nondoxastically accept that what you are doing is right, just as it is one thing to go to the Palace to see the Queen because you believe she is there, but quite another to go to Baker Street to see Sherlock Holmes because you merely nondoxastically accept that he lives there: the former is intelligible (if over-optimistic), whilst the latter is completely silly. Again, though, I think this worry is too quick. On some models of nondoxastic commitment, accepting a moral claim is not much like accepting a fictional claim in the way distinctive of our engagement with fiction. Acting based on a commitment which is of the sort we have concerning Sherlock Holmes would, indeed, be rationally unintelligible. But acting based on some moral principle which I accept because, for example, I value being the sort of person who (as it seems to me) is bound to accept that principle is perfectly rationally intelligible: my *acting on* that principle is just as important to my own conception of what is required to be the sort of person I want to be as my *accepting* it is, and to fail to act in accordance with it would, from my point of view, be a failure. Thus, our acting on our nondoxastic moral commitments is not only rationally intelligible but also psychologically plausible, for our self-conception and our concern for living up to our ideals are surely plausible psychological motors. The mistake made by those who doubt the motivational efficacy of nondoxastic

[16] The fact that I use examples involving fiction should not distract from the fact that I am *not* thinking of nondoxastic *moral* commitment as involving quite the same sorts of attitudes as we have towards fiction. I use fiction examples because they are relatively uncontroversial. I am thinking of nondoxastic moral commitment as irreducible to other types of nondoxastic commitment, but I must use examples of different nondoxastic attitudes if I am to avoid simply begging the question. The point of using the examples, of course, is that they are *relevantly similar*.

moral commitments is to assume, without justification, that those commitments must be like the relatively (and *only* relatively) motivationally inert commitments we have concerning fictional characters.

So, I take it that we are entitled to assume that nondoxastic moral commitments are typically better than no moral commitments at all, at least to the extent that they are *good* moral commitments and not bad or misguided ones.

It is not only moral realists who are entitled to say that there is a way we ought to act and that moral beliefs might not be up to the task of getting us to act that way when it matters most. Non-realists are entitled to distinguish good moral commitments from bad ones, if only on the basis of the *instrumental* non-moral goods associated with some moral commitments as opposed to others. So, non-realists are able to avail themselves of the argument I am giving just as much as realists are. But we can also interpret the talk of goodness in this argument as referring to *moral* goodness, if we are realists. And doing so might well strengthen the argument, if moral goods are particularly important, as a realist might well maintain.

Bringing the strands together, the argument is as follows. It is morally better to have some good moral commitments rather than none. But it would be morally best in at least one respect if those commitments were nondoxastic. The argument for that claim goes like this:

1. Subjective warrant for *beliefs* depends upon one's confidence in the *epistemic* grounds one has;
2. If our moral commitments amount to *beliefs*, then in at least some circumstances we will fail to act on our moral commitments (including our *good* moral commitments, acting on which is acting *well*) if we lack confidence in our epistemic grounds [from our discussion of subjective warrant and action, and (1)];
3. In at least some circumstances, some of us *do* lack confidence in our epistemic grounds for moral beliefs [from our discussion of modesty and scepticism];
4. So [from (2) and (3)] if our moral commitments amount to *beliefs*, then in at least some circumstances we will fail to act on our good moral commitments, i.e. *we will fail to act well*.
5. Subjective warrant for *nondoxastic commitments* does **not** depend upon one's confidence in the *epistemic* grounds one has;
6. So [I take it this follows from (5)], if our moral commitments are *nondoxastic*, then lack of confidence in our *epistemic* grounds need not undermine our fully upholding them in any circumstances, i.e. need not mean that we fail to *act well*.
7. Therefore [from (4) and (6)], it is morally better – in so far as acting well is concerned – if our moral commitments are *nondoxastic* rather than amounting to moral *beliefs*.

This argument is concerned only with our *good* moral commitments. It might turn out that when thinking about our *bad* moral commitments it seems much better that moral commitments are doxastic rather than nondoxastic. Because I have not addressed this issue, I do not take myself to have presented an argument to the conclusion that our moral commitments *ought* to be nondoxastic. My conclusion is evaluative and not normative precisely because the good – and especially the good *in some respect* – does not entail the right, and indeed the good in some respect does not even entail the all-things-considered good. We should take bad moral commitments – doxastic and nondoxastic – very seriously, and I will not try to settle any issues surrounding them here. But given that belief can be dogmatic and hyper-enkratic, it is certainly not clear that bad moral beliefs are less morally dangerous that bad nondoxastic moral commitments.

Thinking it good that our moral commitments be nondoxastic does not by any means commit us to thinking it good that our moral commitments are somehow immune to debate, reason and revision. There is no reason to think that nondoxastic moral commitments are *groundless*, and if they are admitted to have grounds it makes sense to debate, learn about, revise and improve them. If nondoxastic moral commitment is more choice-like than belief-like, that only commits us to thinking of such commitments as immune from rational reflection if we think of choice and preference as immune from such reflection. In fact, it seems that we can and do reflect upon and debate the appropriateness of choices and preferences, without thinking that those choices and preferences are disguised beliefs.[17]

Some philosophers do not understand why I think that realists should consider endorsing nondoxastic acceptance as a response to the moral dangers of sub-enkrasia. They acknowledge that it might well be morally bad that people do not act on their moral commitments, when those commitments are morally good. But since a realist thinks that those commitments are morally good in virtue of expressing moral *truths*, they could simply point out that if someone lacks subjective warrant for *belief* in those moral truths that person ought to inquire further to knowingly acquire sufficient objective warrant. (This might not be a matter of acquiring new or improved grounds for their belief as such, so much as uncovering the mistake in the sceptical argument they were convinced by, or coming to see that epistemic modesty requires less than they thought.) For a realist, according to this line of thought, the natural response to (unjustified) loss of subjective warrant is a call for further or better inquiry, *not* a call to give up on moral beliefs and embrace nondoxastic commitment instead.

Some go even further: it is *morally bad*, they say, to act on commitments which fall short of (well grounded) moral beliefs. If a person does the right thing but does it based on some commitment which is not a belief, they do the right thing for the *wrong reason*, and whilst it might be good that they do the right thing, it would be

[17] It is crucial to recall that this way of putting things does not mean that the fictionalist is embracing any sort of *expressivism*: expressivists make a great deal of how choice- or preference-like moral attitudes are, and they also point out the degree to which our choices and preferences are not beyond the purview of rational reflection. But they do this in support of a *semantic* thesis about the *content* of moral attitudes and language, and the fictionalist rejects this semantic thesis in favour of a standard *representationalist* semantics. See Kalderon (2005a, 2008).

better if they did it for the right reason. So, the realistic fictionalism I have tried to motivate is really a council of moral despair.[18]

Well, sometimes the moral situation is desperate; and when it is, despair seems like an appropriate response and a council of despair the most realistic advice. That makes things sound worse than I think they are, but the idea that what realists should do is *just* recommend or require further enquiry unto confidence in the grounds of one's moral beliefs strikes me as Pollyanna-ish. As I said in the Introduction, the argument I am presenting is to do with the ethics of commitment, and ethics should be sensitive to how things are. Asking, motivating and cajoling people into further enquiry will not, as a matter of fact, free them from the sceptical and modest doubts about the grounds of their moral beliefs which I adverted to above. So even if it is right that acting based on (true) well-grounded moral beliefs is better than acting based on (true) nondoxastic moral commitments, the best is made the enemy of the good if we say that only true, well-grounded moral beliefs are worth aiming for (and not true nondoxastic moral commitments).

Brainwashing might do the trick: widespread inculcation of beliefs which everybody is brought up to think of as indubitably supported by all the available evidence, and active discouragement of reflective inquiry into the grounds of those beliefs, might more or less eliminate the sorts of threats to subjective warrant that I adverted to. But if that is how widespread subjective warrant sufficient to eliminate sub-enkrasia is achieved, it is surely not worth having.[19] I'm sure nobody resistant to my argument on the grounds suggested above would say that brainwashing is what the moral realist should endorse. But morality being as it is, serious reflection breeds scepticism and modesty-induced sub-enkrasia, so they make the best the enemy of the good if they insist always upon open reflection and nothing short of action on the basis of beliefs. How much more *morally* attractive is the alternative I have proposed, in which moral commitments are not constrained evidentially, but are adopted on serious grounds which are open to reflection and challenge and which, perhaps, amount to choices about the sorts of people we want to be?

In any case, it is questionable whether doing the right thing *for the right reasons* requires acting based on beliefs, however well supported they are. As Arpaly (2002: Ch. 3) has shown, there is a perfectly intelligible characterisation of moral worth which does not require that a person has any beliefs about what makes their action right, or even about the rightness of their action. It might, even according to this characterisation, be *best* to act under the motivation of well-grounded (or even not well-grounded) moral beliefs, in which case my point above about the best being the enemy of the good comes into play. But it might well be that according to this or a similar characterisation it is not even better to act under such motivation, as opposed

[18]This objection was pressed by an anonymous referee, who pointed out quite correctly that Clifford would have almost certainly made this reply to my argument. For a strident Kantian defence of the Cliffordian view underlying this objection, see Wood (2003).

[19]Certainly, nobody sympathetic to the Kantian view about the moral desirability of Cliffordian requirements to have properly grounded beliefs expressed by Wood (2003) would think this sort of subjective warrant for beliefs worth having.

to simply doing the right thing *in response to* the right reasons, where ‘in response to’ is cashed out as not requiring the involvement of any sort of representational mental states, including beliefs (and, of course, nondoxastic moral commitments). Then, the only salient question in the debate about whether beliefs or nondoxastic commitments are best for moral purposes would be about which motivates better, in the epistemic and psychological circumstances. The question of which would be the morally better motivation (over and above the moral goodness of getting us to do the right thing) would not arise here, for *no* sort of belief- or nondoxastic commitment-like mental state would be better than any other (so long as they motivate to the same extent).

But one might worry that the argument I presented is too quick: surely the best thing for a moral realist to recommend, considering the danger of sub-enkrasia I highlighted, is for us to form *plans of action* rather than non-doxastic attitudes of acceptance. I might worry about whether my moral beliefs are well grounded, but so long as I have formulated a plan for myself (to not shield cuckolds from the truth, for example) my adopting that plan will be ample motivation for me to act, regardless of the motivational oomph of my beliefs (about the rights of cuckolds).[20]

It is true that plans will serve to regulate my actions once I have adopted them. (See Bratman (1987) for discussion.) But plans are also apt for revision, and in circumstances where the moral stakes are highest it seems plausible that our resolution to stick to the plan (where the alternative is not to ignore it, but to revise it) depends upon our moral commitments: if I cannot endorse the moral principle or consideration which would justify my going on as planned, I have every reason to revise or abandon that plan, at least if going on as planned risks some moral cost. Therefore, adopting plans or policies for action cannot *replace* adopting first-order moral commitments, and lack of subjective warrant is as threatening to our carrying out good moral plans as it is to acting well in general.

17.5 Conclusion

I hope to have shown that there is an argument for fictionalism – not yet full-blown, normative, *revolutionary* fictionalism which says that we *ought* to make our moral commitments nondoxastic, but a weaker, *evaluative* fictionalism which says that there is at least some important respect in which it would be good to do so (if they are not already) – which is not only compatible with moral realism but which a realist is in the best position to advance. Realists have no need to be embarrassed about saying that some moral commitments are ones we *ought to have*, perhaps because they are *true*. And they perhaps have the most reason to think that it is *morally bad* to have moral commitments which we are unlikely to act on in the most morally crucial situations (when those commitments are the right ones).

[20] I thank an anonymous referee for this suggestion.

I said that this would be an essay on the ethics of nondoxastic acceptance. What I have argued is that it is not necessarily *epistemically* bad in any way to have moral beliefs as opposed to nondoxastic moral commitments, but that it is plausibly *morally* bad. That is a conclusion concerning the ethics of commitment in the most authentic sense, the sense in which Clifford felt it an affront to morality, and not just to epistemic norms, to fail in one's epistemic duty. If what I have said is right, fictionalism can be motivated without relying on any contentious meta-ethical views about the metaphysical or epistemological status of moral facts. It can be motivated from within a first-order moral debate about the ethics of commitment.[21]

Acknowledgments Thanks to the audience at a parallel session of the Joint Sessions of the Aristotelian Society and Mind Association in Exeter, 2013 for comments, and to Mark Kalderon and Ralph Wedgwood for discussion of issues related to this argument.

References

Adams, R. M. (1995). Moral faith. *The Journal of Philosophy, 92*(2), 75–95.
Arpaly, N. (2002). *Unprincipled virtue*. Oxford: Oxford University Press.
Bratman, M. E. (1987). *Intentions, plans, and practical reason.* Centre for the Study of Language and Information.
Bratman, M. E. (1992). Practical reasoning and acceptance in a context. *Mind (New Series), 101*(401), 1–15.
Brock, S. (2002). Fictionalism about fictional characters. *Noûs, 36*, 1.
W. K. Clifford, 1876–7 The ethics of belief, Contemporary Review 29, 289.
Coates, A. (2012). Rational epistemic Akrasia. *American Philosophical Quarterly, 49*(2), 113–124.
Cohen, G. A. (2000). *If you're an Egalitarian, how come you're so rich?* Cambridge, MA: Harvard University Press.
Field, H. (1980). *Science without numbers: A defense of nominalism*. Oxford: Blackwell.
Field, H. (1989). *Realism, mathematics & modality*. Oxford: Blackwell.
Harcourt, E. (Ed.). (2000). *Morality, reflection, and ideology*. Oxford: Oxford University Press.
Harman, G. (1986). *Change in view*. Cambridge, MA: MIT Press.
Harman, G. (1999). Rationality. In *Reasoning, meaning, and mind*. Oxford: Oxford University Press
Jay, C. (2012). *Realistic fictionalism.* UCL PhD thesis, available at: http://discovery.ucl.ac.uk/1356103/
Jay, C. (2014). The Kantian moral hazard argument for religious fictionalism. *International Journal for Philosophy of Religion, 75*(3), 207–232.
Jay, C. (2016). Testimony, belief and nondoxastic faith: The Humean argument for religious fictionalism. *Religious Studies, 52*(2), 247–261.

[21] Just as non-realist fictionalists employ (non-moral) normative or evaluative claims to complete their arguments for their views (e.g. Joyce appeals to *prudential* goodness), I employ a moral claim to complete mine. Without such appeals to normative or evaluative claims, neither I nor any other evaluative or revolutionary fictionalist could argue not only that one *might* accept but not believe moral claims, but that it would be good to do so. In this respect, then, my argument is novel in so far as it appeals to *moral* claims, but it no more elides the distinction between first-order ethics and meta-ethics than Joyce's argument elides the distinction between prudence and meta-ethics.

Joyce, R. (2001). *The myth of morality*. Cambridge: Cambridge University Press.
Joyce, R. (2005). Moral fictionalism. In Kalderon (2005b).
Joyce, R. (2006). *The evolution of morality*. Cambridge, MA: MIT Press.
Kahane, G. (2011). Evolutionary debunking arguments. *Noûs, 45*(1), 103–125.
Kalderon, M. E. (2005a). *Moral fictionalism*. Oxford: Oxford University Press.
Kalderon, M. E. (Ed.). (2005b). *Fictionalism in metaphysics*. Oxford: Oxford University Press.
Kalderon, M. E. (2008). Summary [of moral fictionalism]. *Philosophical Books*, *49*(1), 1–3.
Kroon, F. (2011). Fictionalism in metaphysics. *Philosophy Compass, 6*(11), 786–803.
Leng, M. (2010). *Mathematics and reality*. Oxford: Oxford University Press.
Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. Harmondsworth: Penguin.
Friedrich Nietzsche, (1887, 1967). On the genealogy of morals. In W. Kaufmann (trans.), *On the genealogy of morals and Ecce Homo*. New York: Random House.
Nolan, D. (2005). Fictionalist attitudes about fictional matters. In Kalderon (2005b).
Rosen, G. (1990). Modal fictionalism. *Mind, 99*(395), 327–354.
Roush, S. (2017). Epistemic self-doubt. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2017 ed.). https://plato.stanford.edu/archives/win2017/entries/epistemic-self-doubt/.
Sinhababu, N. (2007). Vengeful thinking and moral epistemology. In B. Leiter & N. Sinhababu (Eds.), *Nietzsche and morality*. Oxford: Oxford University Press.
Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies, 27*(1), 109–166.
van Fraassen, B. C. (1980). *The scientific image*. Oxford: Oxford University Press.
Velleman, J. D. (2000). On the aim of belief. In *The possibility of practical reason*. Oxford: Clarendon Press.
Weatherson, B. (2008). Deontology and Descartes' demon. *Journal of Philosophy, 105*(9), 540–569.
Wedgwood, R. (2011). Justified inference. *Synthese, 189*(2), 1–23.
Williams, B. (1975, 1981). The truth in relativism, reprinted. In *Moral luck.* Cambridge: Cambridge University Press.
Williamson, T. (2011). Improbable knowing. In T. Doherty (Ed.), *Evidentialism and its discontents*. Oxford: Oxford University Press.
Wood, A. W. (2003). *Unsettling obligations: Essays on reason, reality and the ethics of belief.* Chicago: University of Chicago Press.
Yablo, S. (2002). Go figure: A path through fictionalism. *Midwest Studies in Philosophy, 25*, 72–102.
Yablo, S. (2005). The myth of the seven. In Kalderon (2005b).