2017, 2 (2)

# ARGUMENTA

The Journal of the Italian Society for Analytic Philosophy

First published 2017 by University of Sassari



Produced and designed for digital publication by the *Argumenta* Staff





*Argumenta* is the official journal of the Italian Society for Analytic Philosophy (SIFA). It was founded in 2014 in response to a common demand for the creation of an Italian journal explicitly devoted to the publication of high quality research in analytic philosophy. From the beginning *Argumenta* was conceived as an international journal, and has benefitted from the cooperation of some of the most distinguished Italian and non-Italian scholars in all areas of analytic philosophy.

Contents

# Editorial

This second issue of the second volume of *Argumenta* opens with the Special Issue *Thinking the (Im)possible*, edited by Carola Barbero, Andrea Iacona and Alberto Voltolini. As the reader will appreciate, it assembles some of the leading researchers on the topic, marking an original contribution to a field of thought which has been a bone of contention throughout the whole history of philosophy, and has been further been excavated at the beginning of the present century—as the Guest Editors make clear in their informative Introduction.

After the Special Issue, we present an article on the philosophy of emotion, written by Talia Morag, who offers both commonsensical and scientific objections to the received view about the nature of the emotions, which she calls "the tracking Dogma". Her challenging theses are very likely to provide substantial fuel for broadening and deepening the discussion of this theme.

Little more than a month ago the international philosophical community suffered from a great loss. Eva Picardi—a philosopher who greatly contributed to the development of analytic philosophy in Italy and abroad—passed away. In his obituary, Paolo Leonardi helps us to understand why and how her role in the contemporary philosophical reflection is irreplaceable.

The section of Book Reviews then rounds off the number. It is the first time that this section has appeared in *Argumenta*, and we are proud to offer readers three dispassionate reviews on as many interesting books.

As usual, all the articles appearing in *Argumenta* are freely accessible and freely downloadable; once again, we are very grateful to the colleagues who have acted as referees.

Buona lettura!

Massimo Dell'Utri

Editor

*Argumenta* 2, 2 (2017): 167
ISSN 2465-2334

*Argumenta* 2,2 (2017)
Special Issue

# Thinking the (Im)possible

Edited by

Carola Barbero, Andrea Iacona and Alberto Voltolini

The Journal of the Italian Society for Analytic Philosophy

Contents

# Introduction: Thinking the (Im)possible

*Carola Barbero, Andrea Iacona, Alberto Voltolini*

*University of Turin*

## 1. The State of the Art

The issue of the relationship between our cogitative abilities, in particular the ability of thinking about something that does not exist, and modal characteristics, in particular those featuring unactualized (im)possibilities, i.e., the ways the world might (not) have been, has always been very intricate.

In analytic philosophy, reflection on this matter has started by reviving an optimistic thesis traditionally ascribed to Hume, according to which conceivability entails possibility: if something is conceivable, then it is also possible. As Wittgenstein clearly suggests in the incipit of the *Tractatus logico-philosophicus*, where he says that that there is no room for conceptions of something impossible:

> The book will, therefore, draw a limit to thinking, or rather—not to thinking, but to the expression of thoughts; for, in order to draw a limit to thinking we should have to be able to think both sides of this limit (we should therefore have to be able to think what cannot be thought). The limit can, therefore, only be drawn in language and what lies on the other side of the limit will be simply nonsense.

On the one hand, the Humean thesis seems quite reasonable. What can be the source of our notion of an unactualized possibility, if not our capacity of conceiving viz. imagining, at least in the sense of imagining which for Hume seemed to be synonymous with conceiving itself (a nonsensuous imagining),[1] how things would have been differently from how they actually are? Yet on the other hand, our imaginative capacities seem well to exceed the realm of the unactualized possible. As Priest (2016[2]) has reminded us with the story of Sylvan's box, which is both empty and full at one and the same time, fiction is plenty of descriptions of situations that might not have obtained.

In (1971, 1980) Kripke seems to have definitely relinquished optimism by drawing a distinction between an *epistemic* notion of possibility, what is possible according to one's state of knowledge, and a *metaphysical* notion of possibility, how things might have really been in the world. For, he says, there are many things that are epistemically possible but not metaphysically possible. For instance, for all what ancients did know, it was surely possible that Hesperus is not identical with Phosphorus. Yet clearly enough, this nonidentity does not amount

[1] Cf. Szabo Gendler and Hawthorne 2002: 17.

*Argumenta* 2, 2 (2017): 173-180
ISSN 2465-2334


DOI 10.23811/40.arg2017.bar.iac.vol

to a metaphysical possibility. Hesperus, viz. Venus, might have been different in many ways from what it actually is—for instance, it might have orbited closer to the Sun—but it might not have been different from Phosphorus, if this means the impossible eventuality for Venus to be different from itself.[2] Granted, adds Kripke, an epistemic possibility corresponds (if not amounts) to a certain sort of metaphysical possibility. In the Hesperus/Phosphorus case, the epistemic possibility that Hesperus is not identical with Phosphorus corresponds to the metaphysical possibility that a given subject be in the same kind of mental states she actually entertains when she faces Venus at dusk and Venus at dawn respectively and yet face different celestial bodies, a certain Hesperean planet and another Phosphorean planet respectively.[3] Yet that metaphysical possibility is *not* the metaphysical possibility that Hesperus is not identical with Phosphorus, for there is no such possibility. As a result, there is a sense according to which what we can conceive—an epistemic possibility—is not a metaphysical possibility, so that the Humean thesis does not hold across the board.

To be sure, one may interpret Kripke's notion of an epistemic possibility as displaying a mere illusion of conceivability. When it seems to us that Hesperus might not have been identical with Phosphorus, it also seems to us that we are conceiving such an (epistemically) possible situation, but in point of fact we are wrong: we are not conceiving that situation, at most we are conceiving another situation that indeed is metaphysically possible, namely the aforementioned situation in which we are in the same mental states as we actually are and yet we face different celestial bodies.[4] If this were the case, clearly enough the entailment between conceivability and possibility could be saved. Yet, even if one may admit that we are not infallible as to what we can conceive, it is hardly the case that this interpretation can be generalized to all cases of seeming conceivability. Sometimes at least, when we seem to conceive something—for instance, when we seem to think that there are no numbers, as an evil demon may lead us into mumbling—we do conceive that very something, even if it is not possible; we are hardly considering the possibility that there are pseudonumbers.[5] Thus, it is better to say, as some have maintained,[6] that an epistemic possibility in general is an illusion of possibility, i.e., the fact that non-P seems possible when in point of fact it is not, for P is instead necessary.

Going in the same direction, following Chalmers one may draw a distinction between a *negative* sense of conceivability, according to which something is conceivable "when it is not ruled out a priori" (Chalmers 2002: 149), and a *positive* sense of conceivability, which, by relying on some (again nonphenomenal, in particular nonsensuous) form of imagination, "require that one can form some sort of positive conception of a situation in which [something] is the case" (*ibid*. 150). The point of such a distinction is precisely to restore a sense of conceivability for

[2] For an alternative account according to which for Hesperus not to be the same as Phosphorus counts as a genuine metaphysical possibility insofar as it does not amount to the impossible eventuality that Venus is different from itself, cf. Voltolini 2014.
[3] Cf. Kripke 1971: 157 fn.15, 1980: 103-104, 143.
[4] Cf. Di Francesco and Tomasetta 2014: 195.
[5] *Ibid*.
[6] Cf. Szabo Gendler and Hawthorne 2002: 33.

which Hume's thesis holds. At most, negative conceivability entails *logical* possibility, the mere noncontradictoriety of a certain imagined situation. Yet it does not entail metaphysical possibility, which is rather what positive conceivability opens the way to. We cannot follow here the details by means of which such a move may be developed, which involve endorsing some form of semantic bidimensionalism, according to which expressions have two forms of intensions, a primary and a secondary one, which in the cases where conceivability really entails metaphysical possibility collapse.[7] Yet even if this move were successful, it remains that there is a residual form of conceivability for which, *pace* Humeans, the entailment to a genuine form of possibility is prevented.

So far, so good, Meinongians of all sorts may say. In actual fact, they hasten to add, even the entailment from a sense of conceivability to logical possibility is to be questioned, for contradictory situations are clearly conceivable. Those who are interested in giving a modal twist to Meinongianism will indeed say that, along with possible states of affairs, hence possible worlds, there also are impossible states of affairs, hence impossible worlds, those where contradictory situations hold. Along with ways the world might have been, there indeed are also ways the world might *not* have been.[8] Those latter ways correspond to what is conceived but it is not possible, for it precisely subsists in some of the latter worlds.[9] To be sure, this move does not please all those who believe that ontology should not be inflated with impossible states of affairs and impossible worlds. One only needs a way to alternatively deal with any evidence that apparently supports the claim that there are impossible worlds. Yet even if one dispenses with impossible worlds and states of affairs, one has still to provide an alternative account of how we can conceive what is not possible.

The situation is made more difficult by the fact that there seem to be not only impossible states of affairs, hence impossible worlds, but also impossible *objects*. Clearly enough, being committed to impossible worlds does not entail being committed to impossible objects. For even if impossible objects would exist at impossible worlds, there may be impossible worlds without impossible objects, namely those worlds containing impossible situations made just by actual, or even merely possible, entities. Yet the intentionality of our thoughts seems to commit us to impossible objects, at least if we admit that there are nonpropositional forms of thoughts. Indeed, the fact that one thinks *of* something, or in other terms, that one's thought has an intentional object, appears not to be exhausted by the fact that one thinks *that* such and such is the case.[10] But if this is the case, then one has thoughts about not only actual, but also about possible and even impossible items, as in the famous example originally pointed out by Twardowski (1894) of someone thinking of a wooden cannon entirely made of steel at one and the same time.

Now, if one wants to dispense with impossible objects, one has to develop an account of intentionality in which one explains how one can think of such objects even if there really are no such objects. This explanation is no easy task.

[7] For such a move, cf. primarily Chalmers 2002.
[8] Cf. Yagisawa 2010.
[9] Cf. Priest 2016[2], Berto 2012.
[10] Cf. e.g. Crane 2001, 2013.

Either one has to resort to a notion of an *intentional content* to be traced back at least to Husserl's *Logical Investigations* (1901), by holding that in thinking of an *impossibile*, one is merely related to an (unsatisfied) intentional content. Or one has to draw a distinction between the ontologically noncommittal notion of an object and the ontologically committal notion of an entity and say that in thinking of an *impossibile*, there is an impossible intentional object one is thinking of even if such an object is no entity.[11] Neither move seems to be unproblematic, unless it is spelled out in appropriate details. What exactly is an intentional content, especially if there really is no object it relates a thinker with (does at least mental descriptivism come back from the rear door)? How can there be impossible intentional objects, if in the overall ontological domain there really are no such things?[12]

These problems appear even more serious if one takes that the domain of objects that necessarily fails to exist is broader than what originally seemed. For there is a sense according to which not only *impossibilia* like the Twardowskian wooden cannon made of steel, but also fictional objects like Madame Bovary and Sherlock Holmes, if not all abstract objects in general (universal attributes among them), necessarily fail to exist.[13]

This issue is intended to reconsider all these venerable problems and indicate possible solutions to them. We hope that these essays will advance the current debate about possibility and impossibility (as well as their conceptions), which seems to be of interest to an increasing number of researchers in various areas such as logic, philosophy of language, epistemology, philosophy of mind, and metaphysics.[14]

## 2. Summaries of Papers

G. Priest, *Thinking the Impossible*

By acknowledging the essential role played by possibility in Western philosophy, Priest focuses on its actual role in possible-world semantics and explains the move from a mono-modal logic to a multimodal logic (whose language contains multiple possibility/necessity operators, and whose semantics contains multiple accessibility relations). There is a most general notion of possibility (simply called "logical possibility") that needs to be taken into account according to which to be possible is to hold at some world, and anything that is possible in any more restricted sense should be possible in this sense. Priest concentrates next on the relation between impossible worlds and possible world semantics by examining two

[11] For the second move see Smith 2002, Crane 2001, 2013, Sainsbury 2010, Sainsbury and Tye 2012. Sainsbury (this issue) returns to it in great detail. Actually, these two moves are not incompatible.

[12] For some such problems, cf. Voltolini 2016.

[13] On this, see Kripke 2013 and Zalta 1983, 1989 respectively.

[14] "Thinking the (Im)possible" was also the title of the FINO/SIFA Graduate Conference that took place in Turin on June 29-30, 2015. On that occasion we started discussing on these problems with some prestigious experts such as M. Sainsbury, G. Priest and T. Williamson together with many researchers and students. After those fruitful discussions we got to like this topic more and more, which prompted us to edit an issue of *Argumenta* on these themes.

different directives, the primary ("Everything holds at some worlds, and everything fails at some worlds") and the secondary ("If A and B are distinct formulas, there are worlds where A holds and B fails") one. The final part of the paper is dedicated to the comparison between conceivability and possibility, where *conceiving* (considered as the same as *imagining*) is seen—differently from the Humean proposal according to which what is conceivable is somehow possible—as the mere bringing before the mind of a particular state of affairs, even of an impossible one. Then he answers three possible objections against the idea of conceiving impossibilities and concludes that the impossible shouldn't be marginalized but understood in its big potential.

T. Williamson, *Counterpossibles in Semantics and Metaphysics*
The paper focuses on counterpossibles, which are counterfactual conditionals with impossible antecedents such as "If whales were fish, their behaviour would differ from what it actually is" or "If whales were fish, their behaviour would be just as it actually is". According to semantic orthodoxy all counterpossibles are true, therefore the above two counterpossibles should be seen as true, and not as genuine alternatives as they seem. Williamson asks whether orthodoxy about counterpossibles is correct and, by defending orthodoxy against recent objections, shows how that kind of questions, far from being a negligible small point of the logic and semantics of counterfactuals, has important ramifications in several directions.

L. Estrada-González, *Impossible Worlds and the Intensional Sense of 'And'*
The essay shows why in an argument like that offered by Lewis against concrete impossible worlds the extensionality of 'and' is an assumption that can be coherently challenged and rejected. Estrada-Gonzales starts by explaining why, independently from Lewis' argument, 'and' is in general intensional and then answers some possible objections. Later on he presents an allegedly 'and'-free argument against impossible worlds which is still subject to well-known objections to the extensionality of 'not'. Then he shows that the reasons to support the intensional 'not' blocking that argument belong to the same family of reasons to support the intensional 'and'. Estrada-Gonzales concludes by claiming that intensional 'and' is needed as a premise-binder and that the argument is blocked at a stage prior to the steps about negation.

T. Yagisawa, *S4 to 5D*
The paper focuses on the modal logic axiom 4, *if necessarily P, then necessarily necessarily P*. According to Chandler (1976) and Salmon (1981, 1989), there is trouble with S4. Axiom 4 is equivalent to *if possibly possibly p, then possibly p* which requires that the accessibility relation between worlds be transitive. Chandler's and Salmon's argument against axiom 4 is based on the idea that even if an ordinary object could have had a slightly different origin from the one it actually has, it could not have had a very different origin from its actual one. Hence, they conclude that accessibility is not transitive, i.e., that what is possibly possible may not be possible. Yagisawa's move is to propose a different way to save axiom 4: by supporting five-dimensionalism, he preserves both axiom 4 and absolute possibility by postulating objects as extended not only in physical space-time but in logical space as well.

V. Morato, *World Stories and Maximality*
The paper deals with the actualist conceptions of modality that reduce talk about possible worlds to talk about world stories. Such conceptions have classical problems, namely that of representing the possible existence of non-actual objects, and that of expressing, in an actualistic way, the possible nonexistence of actual objects. Morato finds a way out of problems of this sort. He suggests that we abandon the notion of global maximality in favor of the notion of local maximality, thanks to which we could generate world stories where the possible nonexistence of an object is represented by the lack of any proposition having it as a constituent. Such world stories would also be locally maximal in the sense of being complete descriptions of alternative courses of actuality.

C. Nencha, *Natural Properties Do Not Support Essentialism in Counterpart Theory: A Reflection on Buras's Proposal*
The paper is a defence of Lewis' antiessentialism against Buras' view (2006). According to Buras, if Lewis accepts both counterpart theory and natural properties, then he can no longer be an antiessentialist: for natural properties determine the existence of similarity relations among individuals that are relevant independently of the ways those individuals are represented, therefore individuals do have real essential properties. Nencha's argument is that the implications of counterpart theory for essentialism are not altered by the acknowledgement of natural properties, since if counterpart theory is antiessentialist without natural properties, then it remains so also when natural properties are taken into account.

M. Jago, *Propositions as Truthmaker Conditions*
The paper outlines an account of propositions as sets of truthmakers, along the lines suggested by Fine (2014 a, b, 2016). According to Jago, propositions are to be seen as sets of possible truthmakers, thanks to which he succeeds in offering a redescription of semantic phenomena such as same-saying, subject matter, and aboutness.

N. Spinelli, *Husserlian Intentionality and Contingent Universals*
Spinelli starts from Husserl's challenge of maintaining both that universals exist in the strongest sense and that they exist contingently. After a short presentation of Husserl's intentionalism, idealism and the role played by universals, he then presents a version of the Husserlian view regimented in terms of modal logic and possible-worlds semantics and distinguishes between two accessibility relations, world-bound and free, having different structural properties. Thanks to his modal apparatus, he is able to show how the necessary or the contingent existence of universals can be derived. Therefore, he concludes that in Husserl's philosophy there is room for both necessary and contingent universals.

M. Sainsbury, *Intentional Relations*
The paper focuses on a classical topic concerning intentionality that could be summed up by the following question: what kind of relation is the intentional relation? Is it a two-term relation or a three-term relation? Sainsbury starts from the intuition that, on the one hand, thinking about Obama and thinking about Pegasus seem to be the same kind of thing (since both are cases of thinking about something), but, on the other hand, they also seem to be different kinds of thing because the first kind of thinking seems to be relational whereas the other does

not. The essay aims at offering a solution to this kind of problems by distinguishing varieties of relationality and by underlining that what matters is the two-term relational nature of all intentional states, regardless of whether or not the representations they involve have referents.

## References

Berto, F. 2012, *Existence as a Real Property*, Dordrecht: Springer.

Buras, T. 2006, "Counterpart Theory, Natural Properties and Essentialism", *Journal of Philosophy*, 103, 27-42.

Chalmers, D. 2002, "Does Conceivability Entail Possibility?", in Szabo Gendler, T. and Hawthorne, J. (eds.), *Conceivability and Possibility*, Oxford: Oxford University Press, 145-200.

Chandler, H.S. 1976, "Plantinga and the Contingently Possible", *Analysis*, 36: 106-109.

Crane, T. 2001, *Elements of Mind*, Oxford: Oxford University Press.

Crane, T. 2013, *The Objects of Thought*, Oxford: Oxford University Press.

Di Francesco, M. and Tomasetta, A. 2014, "Immaginare e sperimentare. Gli zombie e il problema della coscienza fenomenica", *Rivista di estetica*, 56, 179-208.

Fine, K. 2014a, "A Theory of Truth-Conditional Content I: Conjunction, Disjunction and Negation", *Unpublished manuscript*.

Fine, K. 2014b, "A Theory of Truth-Conditional Content II: Subject-matter, Common Content, Remainder and Ground", *Unpublished manuscript*.

Fine, K. 2016, "Angelic Content", *Journal of Philosophical Logic*, 45, 199-226.

Husserl, E. 1901, *Logical Investigations*, London: Routledge, 2001.

Kripke, S. 1971, "Identity and Necessity", in Munitz, M.K. (ed.), *Identity and Individuation*, New York: New York University Press, 135-64.

Kripke, S. 1980, *Naming and Necessity*, Oxford: Blackwell.

Kripke, S. 2013, *Reference and Existence*, Oxford: Oxford University Press.

Priest, G. 2016[2], *Towards Non-Being*, Oxford: Clarendon Press.

Sainsbury, M. 2010, "Intentionality Without Exotica", in Jeshion, R. (ed.), *New Essays on Singular Thoughts*, Oxford: Oxford University Press, 300-17.

Sainsbury, M. and Tye, M. 2012, *Seven Puzzles of Thought and How to Solve Them*, Oxford: Oxford University Press.

Salmon, N. 1981, *Reference and Essence*, Princeton, NJ: Princeton University Press.

Salmon, N. 1989, "The Logic of What Might Have Been", *The Philosophical Review*, 98, 3-34.

Smith, A.D. 2002, *The Problem of Perception*, Cambridge (MA): Harvard University Press.

Szabo Gendler, T. and Hawthorne, J. 2002, "Introduction: Conceivability and Possibility", in Szabo Gendler, T. and Hawthorne, J. (eds.), *Conceivability and Possibility*, Oxford: Oxford University Press, 1-70.

Twardowski, K. 1894, *Zur Lehre vom Inhalt und Gegenstand der Vorstellungen*, Munich: Philosophia, 1982.

Voltolini, A. 2014, "Contingent Sameness and Necessary Identity", in Palma, A. (ed.), *Castañeda and His Guises*, Berlin: De Gruyter, 197-211.

Voltolini, A. 2016, "Tim Crane, The Objects of Thought", *Dialectica*, 70, 245-52.

Wittgenstein, L. 2012, *Tractatus Logico-Philosophicus*, London: Routledge.

Yagisawa, T. 2010, *Worlds and Individuals, Possible and Otherwise*, Oxford: Oxford University Press.

Zalta, E.N. 1983, *Abstract Objects*, Dordrecht: Reidel.

Zalta, E.N. 1988, *Intensional Logic and the Metaphysics of Intentionality*, Cambridge (MA): The MIT Press.

# Thinking the Impossible

*Graham Priest*

*City University of New York*
*University of Melbourne*

*Abstract*

The article looks at the structure of impossible worlds, and their deployment in the analysis of some intentional notions. In particular, it is argued that one can, in fact, conceive anything, whether or not it is impossible. Thus a semantics of conceivability requires impossible worlds.

*Keywords*: Possible World, Impossible World, Conceivability, David Hume, FDE.

«*Possible? Is anything impossible? Read the newspapers*».
Arthur Wellesley (Duke of Wellington).[1]

## 1. Introduction: The History of Modality

Possibility has been a familiar character in Western philosophy since the inception of the discipline. Systematic ways of thinking about it are to be found in both of the first two great periods of logic: Ancient and Medieval. Witness Aristotle's modal syllogistic[2] and and its medieval developments, such as the doctrine of ampliation, and the notions of *sensu composito* and *sensu diviso*.[3]

The articulation of the notion has taken a very distinctive turn in the third great (and contemporary) period. Possible-world semantics has come to take centre stage. And the applicability of these has spread the tentacles of modality into areas where connections had not before been made, such as meaning, belief, conditionals, and intentionality.[4] Of late, we have seen an extension of this logical technology into the area of impossible worlds, stretching the tentacles further, and—arguably—untangling some knots in the earlier tentacles.[5]

The present paper takes a close look at the structure of impossible worlds, and of one in particular of the tentacles: the mental state of conception/imagination.

[1] Cohen and Cohen 1992: 450.
[2] See Smith 2011.
[3] See Knuuttila 2014.
[4] See Garson 2014.
[5] See Berto 2014.

*Argumenta* 2, 2 (2017): 181-194
ISSN 2465-2334


DOI 10.23811/41.arg2017.pri

# 2. Possible Worlds

## 2.1 Their Structure

Possible-world semantics are familiar to contemporary logicians and philosophers, and need little explanation;[6] but let me set things up in a slightly unusual way, for reasons that will become clear later. Take a propositional language with the connectives: $\wedge$, $\vee$, $\neg$, $\Diamond$, $\Box$.[7] ($\supset$ can be defined in the usual way.) An interpretation for the language has four components: $\langle X, R, @, v\rangle$. $X$ is a set of (possible) worlds. (It would be more normal to write this as $W$; but I will hold this letter in reserve till later.) $R$ is a binary relation on $X$: relative possibility. $@ \in X$ is the actual world. And $v$ assigns every propositional parameter a *pair* of subsets of $X$, $v^+(p)$ and $v^-(p)$, subject to the constraints of exclusivity and exhaustivity:

**Exc**: $v^+(p) \cap v^-(p) = \varnothing$

**Exh**: $v^+(p) \cup v^-(p) = X$

Intuitively, $v^+(p)$ is the set of worlds where $p$ is true; $v^-(p)$ is the set of worlds where $p$ is false.

We now define what it is for a formula to be true ($\Vdash^+$) and false ($\Vdash^-$) at a world $w \in X$:

- $w \Vdash^+ p$ iff $w \in v^+(p)$
- $w \Vdash^- p$ iff $w \in v^-(p)$
- $w \Vdash^+ \neg A$ iff $w \Vdash^- A$
- $w \Vdash^- \neg A$ iff $w \Vdash^+ A$
- $w \Vdash^+ A \wedge B$ iff $w \Vdash^+ A$ and $w \Vdash^+ B$◇
- $w \Vdash^- A \wedge B$ iff $w \Vdash^- A$ or $w \Vdash^- B$
- $w \Vdash^+ A \vee B$ iff $w \Vdash^+ A$ or $w \Vdash^+ B$
- $w \Vdash^- A \vee B$ iff $w \Vdash^- A$ and $w \Vdash^- B$
- $w \Vdash^+ \Diamond A$ iff for some $w'$ such that $wRw'$, $w \Vdash^+ A$
- $w \Vdash^- \Diamond A$ iff for all $w'$ such that $wRw'$, $w \Vdash^- A$
- $w \Vdash^+ \Box A$ iff for all $w'$ such that $wRw'$, $w \Vdash^+ A$
- $w \Vdash^- \Box$ A iff for some $w'$ such that $wRw'$, $w \Vdash^-$ A

Validity ($\vDash$) is defined as preservation of truth at @ in every interpretation.[8]

A simple induction shows that for every formula, $A$, and world, $w$, $w \Vdash^+ A$ or $w \Vdash^- A$, but not both. Hence, given that no constraints are placed on $R$, the logic delivered is simply the modal logic $K$.∀

It should be noted that an interpretation is simply a piece of mathematical machinery. In particular, $X$ is any old set of objects. These are not to be confused with possible worlds themselves. We may naturally suppose, however,

[6] See Priest 2008: Chs. 2, 3.

[7] What follows applies equally to first-order languages, but their specificities are not relevant to the following considerations.

[8] In some standard presentations, there is no designated world, @, and validity is defined as truth preservation over all possible worlds. As long as no special constraints are put on @, this is, of course, equivalent. However, it will be useful in what follows to have @ at our disposal.

that there is one interpretation of the language which is in accord with the real. (In this, $X$ is the set of real possible worlds, @ is the real actual world, $R$ is the real relation of relative possibility, and $v^+(p)/v^-(p)$ are the sets of worlds in which $p$—understood as some meaningful sentence—is really true/false.) That is why we can reason using modal logic about reality (not just actuality: actuality is just one world of the plurality of worlds). Why do we require a plurality of interpretations to define validity? For the same reason that we do in the case of propositional non-modal logic. We want our inference relation to be applicable whatever reality is, in fact, like.

Of course, none of this tells us what kind of thing possible worlds are. Physical objects or abstract objects, existent objects or non-existent objects? There are many well-known accounts of this matter;[9] and which is right need not concern us in this essay. Of more concern here is possibility itself.

## 2.2 A Plurality of Possibilities

Possibility comes in many flavours. To name but a few: physical ($\varphi$), epistemic ($\varepsilon$), deontic ($\delta$). Let us write $K$ for the set of different kinds of possibility. For every member of $K$ there will be a corresponding notion of necessity. If $\kappa \in K$, let us write the corresponding modal operators as $\langle\kappa\rangle$, and $[\kappa]$. In an interpretation, each $\kappa \in K$ will also have its own accessibility relation, $R_\kappa$. Thus we will have for each $\kappa$:

- $w \Vdash^+ \langle\kappa\rangle A$ iff for some $w'$ such that $wR_\kappa w'$, $w \Vdash^+ A$
- $w \Vdash^- \langle\kappa\rangle A$ iff for all $w'$ such that $wR_\kappa w'$, $w \Vdash^- A\Diamond$
- $w \Vdash^+ [\kappa]A$ iff for all $w'$ such that $wR_\kappa w'$, $w \Vdash^+ A$
- w $\Vdash^- [\kappa]A$ iff for some $w'$ such that $wR_\kappa w'$, $w \Vdash^- A$

What we have now done is to move from a mono-modal logic to a multi-modal logic, whose language contains a multiplicity of possibility/necessity operators, and whose semantics contains a corresponding multiplicity of accessibility relations.[10]

The accessibility relations will come with appropriate constraints. Thus, one would expect that for every world, $w$, $wR_\varepsilon w$, so that $[\varepsilon]A \vDash A$ (what is known is true). Sometimes, the accessibility relations will be nested, in the sense that possibility in one sense implies possibility in another. (We will have an example of this in a moment.) Sometimes there is no nesting. For example, epistemic and physical possibility properly overlap. Thus, there is a physical limit to how fast a marathon can be run. Suppose, for the sake of argument, that this is one hour. Then it is physically possible to run a marathon in 61 minutes, and physically impossible to run a marathon in 59 minutes. But both of these are (currently) epistemic possibilities. Conversely, it is both physically possible and epistemically possible for something to be made of antimatter. In the 13th Century, it was still a physical possibility, but it was not an epistemic possibility: people then had no conception of antimatter, or, therefore, of its possibilities.

For the most part, we will not be concerned with the constraints on the accessibility relations—with one major exception. Given possible-world semantics, there is a most general notion of possibility. To be possible in this sense is

[9] See Menzel 2013. My own account can be found in Priest 2005: 7.3.
[10] See Carnielli and Pizzi 2008.

simply to hold at *some* world. How to describe this kind of modality, one might argue about. I shall simply call it logical, $\lambda$. (Though it is worth noting that logical necessity in this sense will include things that are not formally logically necessary, including analytic truths such as 'all red things are coloured' and mathematical truths such as 'there is an infinitude of prime numbers'.) Anything that is possible in any more restricted sense is possible in this sense; and $R_\lambda$ is simply the universal relation: every possible world accesses every other. Hence, for any $\kappa \in K$, we have:

- $\langle\kappa\rangle A \vDash \langle\lambda\rangle A$
- $[\lambda]A \vDash [\kappa]A$

and the modal logic of $\lambda$ is S5.

Here we see the beginning of a problem. Some things that are epistemically possible would seem to be logically impossible. Thus, before Wiles' proof of the truth of Fermat's last theorem, its negation was epistemically possible, though logically impossible.

## 3. Impossible Worlds

### 3.1 The Primary Directive

The rediscovery of modal logic in the 20th century was in the work of C.I. Lewis between the two World Wars. Possible-world semantics came to prominence in the 1960s and 70s. At first, under the influence of Quine's attack on things modal, possible worlds and their machinations were considered creatures of darkness. But the clarity of the mathematics involved, and their usefulness in an analysis of many things other than modality—such as conditionals, meaning, knowledge and belief—meant that they soon became part of the intellectual landscape. The philosophical debate around worlds changed from *whether* one can make sense of them to *how* best to make sense of them, given the slew of theories about their nature.

Impossible worlds rose to prominence some 20 years later. Under a very different ideology (that of the unintelligibility of inconsistency), they, too, were often taken to be creatures of darkness. Current debates may still concern whether one can make sense of them. However, their mathematics is clear, and their applicability to many philosophical areas—including some of those in which possible worlds were clearly problematic—have, I think, ensured, that they will soon be as much part of the landscape as possible worlds.[11]

Of course, for certain notions of possibility, impossible worlds can be accommodated in possible-world semantics. Thus, physically impossible worlds, where, say, a particle accelerates through the speed of light, are logically possible; and so can be accommodated in a model which allows for all logical possibilities.

The main problem is with logical impossibilities themselves. On standard possible-world semantics there are no worlds which realise these. But if there are worlds which do so, there must be worlds where logical impossibilities hold; and dually, worlds where logical truths fail. There appears to be no reason to distinguish between different kinds of logical truths and falsehoods in this re-

[11] On these issues, see Priest 1997a and Berto 2013.

gard. Hence we have the first leading principle of impossible-world semantics: any logical truth must fail at some worlds, and any logical falsity must hold at some worlds. Of course, this was already the case for logically contingent things. Hence we arrive at what we may call the Primary Directive:

• *Everything holds at some worlds, and everything fails at some worlds.*

What, then, is the technology of logically impossible worlds? We may simply broaden the class of worlds by dropping **Exc** and **Exh**.

It is almost trivial to check that if there are no constraints on accessaibility relations at impossible worlds the Primary Directive is then satisfied. Take any formula, $A$, and consider a world, $w$, such that for all $\kappa \in K$, $w$ accesses itself and only itself under $R_\kappa$, and for every parameter, $p$, in $A$, $w \in v^+(p) \cap v^-(p)$. A simple induction shows that every formula whose parameters are amongst the $p$s—and so $A$—is both true and false at $w$. Dually, replace the condition $w \in v^+(p) \cap v^-(p)$ with $w \notin v^+(p) \cup v^-(p)$, and a similar induction shows that $A$ is neither true nor false at $w$.[12]

## 3.2 Possibile Worlds Revisited

So far so good. Let $P \subseteq X$ be the logically possible worlds. We should clearly require that $@ \in P$. (What is actual is logically possible.) And we should require that the worlds in $P$ access each other, and nothing else, under $R_\lambda$.

But what is $P$? One reason the answer is important is that it determines the validity relation, since this was defined in terms of truth preservation at $@$, and $@$ is in $P$. Those who think that classical logic gets the validity relation right will, of course, suggest that $P$ is a proper subset of $X$, namely, the worlds where **Exh** and **Exc** hold. These are the worlds closed under classical S5 for the modality $\lambda$, except the trivial world (where everything is true).[13]

What if one does not think that classical logic is correct? If one takes *FDE* (First Degree Entailment) to be the correct logic, then we could, at the other extreme, as it were, just take $P$ to be $X$ itself (so there are no impossible worlds—at least of this kind; stay tuned). There are intermediate possibilities. If one takes the correct logic to be *LP*, we will just reinstate the condition **Exh**. The possible worlds are then those that are closed under LP-S5 consequence (including, NB, the trivial world). Alternatively, if one takes the correct logic to be *K3*, we will reinstate the condition **Exc**. The possible worlds are then those closed under *K3*-

[12] I note that if we make all the propositional parameters true and false at $w$, then every formula is true there; and if we make all the propositional parameters neither true nor false at $w$, all formulas are neither true nor false there. Hence, the primary directive can, in fact, be satisfied with just these two worlds. However, the two, on their own, hardly do justice to the diversity of impossible situations.

[13] Any such world is obviously closed under the consequence relation. Conversely, if $w$ does not satisfy **Exh** and **Exc**, it is clearly not closed under the relation. If it accesses a world, $w'$, that is not so closed, then for some, $p$, $p$ is either both true and false at $w'$ or neither true nor false there. In the first case, $\Diamond(p \wedge \neg p)$ is true at $w$; but in S5, $\Diamond(p \wedge \neg p) \vDash A$, so the world is either not closed under the consequence relation, or is trivial. In the second case, it is not true that $\Box(p \vee \neg p)$ at $w$, so the world is not closed under Necessitation.

*S5*, except the trivial world.[14] I note that the Primary Directive may no longer be satisfied for modal formulas involving λ. Thus, in *LP* $\Box(p \vee \neg p)$ will hold at all worlds; and in *K3* $\Diamond(p \wedge \neg p)$ will fail at all worlds. This will be rectified with the Secondary Directive, as we shall see in a moment.

It is sometimes touted as a virtue of possible-world semantics that they provide a reductive account of (logical) possibility. To be possible is simply to hold in *some* world. Exactly the same is true if $P$ is $X$; but of course, it is not true if $P$ is a proper subset of $X$. Reduction has alway struck me as a dubious virtue, however. Why should one expect such a reduction? We obviously don't have it for all the other kinds of possibility; why just this one? Or better, we have to give a non-reductive account for the other notions of possibility, so we ought to be able to do it for this one too. We have just seen how.

More importantly in the present context: a natural thought is that if, at a possible world, something is possible in any sense, it is logically possible. For some notions of possibility this seems right. We would expect any physical possibility to be a logical possibility. And if $X = P$ then everything is logically possible.[15] Indeed, for any $\kappa \in K$, if $w R_\kappa w'$ then $wR_\lambda w'$.

But at least if $P$ is a proper subset of $X$, there are good reasons why this should not hold for all $\kappa \in K$. Consider epistemic possibility. Take a logical untruth, $A$, of enormous complexity: one which it would take longer than the history of the cosmos to decide. As far as is known, $A$ could be true. So there must be some $w \in X - P$, such that $@R_\varepsilon w$.

Or again, suppose, for the sake of illustration, that the Law of Excluded Middle is a logical truth. Let us suppose that intuitionist critiques have been so fierce that we are now no longer sure whether $A$ is true or not, where, this time, $A$ is: either there are or there are not 17 consecutive 0s in the decimal expansion of π. It could be false for all we know; but $A$ is false at no logically possible world, so there must be a $w \in X - P$, such that $@R_\varepsilon w$. Or a more realistic example. Let $G$ be a statement of Goldbach's Conjecture. (Every even number greater than 2 is the sum of two primes.) The conjecture is currently undecided. It is true for all we know; it is false for all we know. Hence there are worlds $w_1$ and $w_2$, such that $G$ is true at $w_1$, false at $w_2$, and $@R_\varepsilon w_1$ and$@R_\varepsilon w_2$. Either $w_1 \in X–P$ or $w_2 \in X–P$.

Or consider deontic modality. I may promise to do something logically impossible (such as prove some mathematical statement which is, as a matter of fact, false). Or I may make promises to do incompatible things, such as to be in two different places at the same time (assuming such to be impossible). I am then morally obliged to do impossible things;[16] that is, the worlds that realize my obligations are impossible. Hence, for any $w$ such that $@R_\delta w$,

[14] For these non-classical modal logics, see Priest 2008: Ch. 11a.

[15] As argued by Mortensen 1989. Even if $X$ is not $P$, this may still be the case. In *FDE* and *LP* everything is logically possible, because of the trivial possible world. The thought that every situation is logically possible may initially seem an odd one. But it should be remembered that logical possibility is a very weak constraint. Even if one is of a classical persuasion, it is a logical possibility that I can jump a kilometre into the air, that the moon is made of blue cheese, etc. Usually, when we are concerned with possibility, we are concerned with much more restricted notions, especially physical possibility.

[16] See Priest 1987, Ch. 13.

$w \in X - P$.

Sometimes, at this point, people will say things like 'Of course we are assuming an ideal agent'. A poor move. It helps us not one iota to understand what it means for us to know or be obliged to do something, if all we understand is what it is for God to know or be obliged to do it. We are interested in notions that apply to us: not God.

### 3.3 The Secondary Directive

The Primary Directive tells us that everything must hold at some world, and everything must fail as some world. There is, however, a stronger condition:

• *If A and B are distinct formulas, there are worlds where A holds and B fails.*

Let us call this the Secondary Directive. (It of course entails the Primary Directive.) The above semantics does not deliver this directive. Thus, for example, if $C$ is true at any $w \in X$, so is $C \vee D$, for any $C$ and $D$. More generally, if $A$ entails $B$ in FDE, and $A$ is true at $w$, so is $B$. Clearly, for some $A$s and $B$s, a world that realises the Secondary Directive must be logically impossible. If one takes as a motto the thought that at an impossible world, anything can happen, the Secondary Directive seems entirely reasonable. Are there stronger reasons?

There are. One concerns conditionals with logically false antecedents. Thus, assuming that intuitionist logic is not the correct logic, the following seem true and false, respectively:

• If intuitionist logic is correct, the Law of Excluded Middle fails.
• If intuitionist logic is correct, the Law of Non-Contradiction fails.

Given something like a standard theory of counterfactuals,[17] to evaluate such conditionals we must consider worlds where intuitionist logic is correct. If $A$ does not entail $B$ in intuitionist logic, there must be such worlds where $A$ holds and $B$ fails. This does not happen in the present semantics. In intuitionist logic $\neg\neg A$ does not entail $A$; but in the present semantics the one holds at a world iff the other does. More generally, consider an arbitrary logic, $L$, and some $A$ which does not entail some $B$, according to $L$. Then to evaluate counterfactuals of the form: 'if $L$ were the correct logic...', we have to consider worlds where $L$ is the correct logic, and so where $A$ may hold and $B$ may fail.[18]

A second reason comes from consideration of intentional states, such as fear, hope, etc.—and crucially in the current context, belief, knowledge, and conception. People being what they are, if $A$ and $B$ are distinct, someone may believe $A$, but not $B$. No one said that we are dealing only with rational agents—whatever that may mean. And, even a rational agent (not an idealised one) may believe/know all the axioms of Peano Arithmetic, without believing/knowing all their consequences. Hence we are required to countenance worlds which realise these states. The Secondary Directive delivers these.

How, technically, are we to obtain states delivered by the Secondary Directive? The simplest way is by brute force. We now augment $X$ with a set of worlds, $Z$, to give us a set of worlds $W = X \cup Z$. At worlds in Z, every formula is treated as atomic. (Priest (2005) calls such worlds open worlds.) Thus, in an

[17] See Priest 2008, Ch. 5.
[18] One notable exception: logics where the inference $A \vdash A$ fails.

interpretation, $v^+/v^-$ applies to arbitrary formulas (not just parameters). The truth conditions at worlds in $X$ are now as before (so if $A$ is not atomic, $v^{\pm}(A)$ is irrelevant). But if $w \in Z$:

- $w \Vdash^+ A$ iff $w \in v^+(A)$
- $w \Vdash^- A$ iff $w \in v^-(A)$

And for $\kappa \in K$, $R_\kappa$ is a binary relation on $W$. The possible worlds are still $P \subseteq X$, and the impossible worlds are $(X-P) \cup Z$. @ is still in $P$. It is easy to see that the Secondary Directive (and so the Primary Directive) is now satisfied.[19]

At this point, one might think one has gone too far. Why does one need to countenance all these worlds? Here is a natural view.[20] Counterfactuals are context dependent. If we consider counterfactuals such as 'If intuitionist logic were correct ...', the context requires us to consider worlds in which intuitionist logic is true; so we need to extend the realm of possibility to those worlds that are intuitionistically possible. Other contexts will require similar extensions. But there is no context in which we need to consider all such worlds. Conditionals are, of course, only one example of topics for which we need impossible worlds. But even concentrating just on conditionals, and granting the analysis given—at least for the sake of argument—even if it is the case that there is no context which requires every world, every world in $Z$ will be required in some context (since we may consider an arbitrary logic). Hence, globally, they must all be available to us. Whether one wants to call the worlds where intuitionist logic holds 'possible' in an extended sense, appears to me to me a terminological matter. Call them so if one wishes; but if intuitionist logic is not the correct logic, they may not be logically possible in the veridical sense.

### 3.4 Generalising

By taking advantage of the fact that the set of truths/falsehoods in any world in $X$ can be imitated by a world in $Z$, we can, in fact, make matters more uniform (and cut out the first stage in the construction of impossible worlds). An interpretation is now a tuple $\langle W, @, \{R_\kappa : \kappa \in K\}, v\rangle$. $W$ is a set of worlds; $@ \in W$; for $\kappa \in K$, $R_\kappa$ is a binary relation on $W$;[21] and for any formula, $A$, $v^{\pm}(A)$ are subsets of $W$. The truth conditions for @ are as in 2.1/2.2, and the truth conditions for any other world, $w$, are as in 3.3. What are the possible worlds? We may simply take these to be those closed under the *S5* version of whichever of our four logics we hold to be correct (or if this is explosive, all such worlds except the trivial world).[22]

The techniques employed here are also generalisable in natural ways to most other standard propositional logics—not just the four we have met: *FDE*,

[19] In truth, it is only $v^+$ that is required to deliver the Secondary Directive. We cannot give up $v^-$, though, since it may be involved in the falsity conditions of modal formulas at @; though this leaves us free to impose constraints on $v^-$ if required for any reason.

[20] Suggested to me by Hartry Field.

[21] In fact, we do not need to consider what is accessed by non-@ worlds under $R_\kappa$, since the truth/falsity of modal sentences at such worlds is taken care of by $v$. We may therefore take $R_\kappa$simply to be of the form $\{\langle @, y\rangle : y \in Y\}$ for some $Y \subseteq W$.

[22] And the new context may suggest some new members of $K$, such as 'It is intuitionistically possible that ...'.

*LP*, *K3*, and classical logic. For example, if we take as our basic propositional logic an LFI,[23] possible worlds have this as their underlying logic, and the negations of modal statements are given non-deterministic truth conditions. Since LFIs are paraconsistent logics, all worlds obtained in this way are possible. Or we may take a propositional logic that itself has world semantics. Thus, the Kripke semantics for intuitionistic logic adds a binary accessibility relation to give the truth conditions of the conditional.[24] All the worlds delivered are possible. Or the Routley/Meyer world semantics of relevant logics add a ternary accessibility relation to give the truth conditions of the conditional.[25] The semantics themselves specify possible (normal) and impossible (non-normal) worlds. In all cases, open worlds may be added to the models, in order to satisfy the Secondary Directive (and so the Primary Directive if it is not already satisfied).

## 4. Conceivability and Possibility

I now want to turn to the notion of conceiving.[26] Perhaps this can be understood in many ways. I intend to use *conceive* here as roughly synonymous with *imagine*: the sort of imagination employed by scientists, mathematicians, philosophers, novelists, political reformers, theologians, visionaries, and so on.[27] In imagination, a state of affairs or an object is brought before the mind, and may be considered, enjoyed, its consequences thought through, and so on. 'Conceive' can be an intentional operator (to conceive *that* something), and it can be an intentional predicate (to conceive an object). Let us start with the intentional operator.

A very traditional view is that if one can conceive of something, it is possible. As David Hume put it:

> 'Tis an establish'd maxim in metaphysics, *That whatever the mind clearly conceives includes the idea of possible existence, or in other words, that nothing we imagine is absolutely impossible* (Hume 1739-40: 32).

Hume's absolute impossibility here is essentially logical impossibility.[28] Hence, for Hume, one cannot conceive of a logical impossibility. In particular, there are things that cannot be conceived.

Even given a Humean conception of what is logically impossible, I have always found this view incredible. Take Goldbach's conjecture again. I have no difficulty in conceiving this, and no trouble conceiving its negation, though one

[23] See Carnielli, Coniglio and Marcos 2007, and Bueno-Soler 2012.
[24] See Priest 2008, Ch. 6.
[25] See Priest 2008, Ch. 10.
[26] *OED*, *to conceive*: 'to take or admit into the mind, to form in the mind, to grasp with the mind'.
[27] OED, *to imagine*: 'to form a mental image of, to represent to oneself in imagination, to create as a mental conception, to conceive'. There is one sense of the word according to which what is imagined 'should not be known with certainty' (*OED*, again). This is not the sense at issue here.
[28] For Hume, for something to be absolutley impossible is for it to imply a contradiction. (See Lightner 1997: 115.) I take it that he holds that the negation of any "relation of ideas" would do this.

of these is mathematically impossible. Indeed, mathematicians must be able to conceive these things, so that they understand what it is of which they are looking for a proof, or so that they can infer things from them, in an attempted *reductio* proof. Nor does the conceivability of Goldbach's conjecture and its negation disappear if I discover which one of them is true, and so the other no longer appears mathematically possible to me. Hence, when something is conceived it may not even be *appear* to be possible.[29]

Similarly, the claim that intuitionist logic is true is, I take it, logically false (and if you disagree, merely replace this with classical logic in the following example). Yet I have no problem in conceiving what it would be like for it to be true. Indeed, I have to do so in order to be able to debate with intuitionists on the matter.

Moreover, I have no problem in imagining that deep in a trench at the bottom of the Pacific Ocean, there is a pearl which is round and square. I cannot form a visual image of this. But imagination should not be confused with visual imagery. I cannot form a visual image of a chiliagon (a regular 1,000 sided figure), even though there is nothing impossible about this. Conversely, I can visually picture a state of stationary motion, even though this is contradictory. This occurs in the well-know "waterfall illusion". In this, one conditions the visual system with constant motion in one direction. The after-image will make things appear to be moving in the opposite direction. But if one focusses on a point in the visual field, it appears to be stationary, even though in motion.[30]

And again, understanding a work of fiction requires an act of imagination. Yet there are works of fiction with essentially inconsistent plots—for example, *Sylvan's Box*.[31] One must therefore be able to imagine such things.◇

Indeed, it seems to me that I can conceive of and imagine *anything* that can be described in terms that I understand.[32] (Which is not to say that such things are the only things I can imagine. That is another matter.) In fact, such understanding allows for the possibility of conception—which is not the same as the conception of possibility. To conceive, I merely have to bring the state of affairs, so described, before the mind.◇

So, to return to the formal semantics, given any of the interpretations with impossible worlds of the kinds described in 3.3, there will be a *really* most general notion of possibility: being true at *some* world. Call this *global possibility,* $\gamma$. If $W$ is the set of all worlds, then for @ (or, more generally, for every world in $P$),

[29] See Yablo 1993. Yablo's own account of conceivability (in Section 10) is that $A$ is conceivable if one can imagine a world that verifies $A$. In fact, I agree with this, since I take everything to be conceivable/imaginable. This is not what Yablo intends, however. For, by 'world', he means '(classically) possible world'. Yablo tells us (30) that one cannot imagine, e.g., tigers that lick all and only those tigers that do not lick themselves. I find this no harder to imagine than a set that contains all those sets which are not members of themselves. (And I could imagine this even before I became a dialetheist.)

[30] For discussion and references, see Priest 2006: 3.3.

[31] Priest 1997b.

[32] There is a somewhat thorny issue here about what it is, exactly, to understand. Can a congenitally blind person understand the predicate 'is red', for example? I am inclined to the view that they can, if they can use the word—by whatever means—in a roughly normal way. When they imagine something red, the phenomenological content may, however, be quite different from that of a sighted person who imagines something red.

$@R\gamma\, w$ for all $w \in W$. Given the Primary Directive, for any $A$, $\langle\gamma\rangle A$ is true at $@$, and $[\gamma]A$ is not true at $@$. We may take $\gamma$ to be the modality of conception/imagination: $\langle\gamma\rangle A$ is '*A* is conceiveable/imaginable'.[33] I should note that, strictly speaking, conceivability is agent-relative (as is knowability). In particular, the *As* in question have to come from a language that the agent in question understands (in a way that, say, a medieval monk could not understand the language of quantum mechanics).[34]

Three objections. *One*. It might be suggested that if I seem to conceive of (imagine) something that is impossible, I am, in fact, conceiving something else. Thus (assuming that identities are necessary), when I conceive that water is not $H_2O$, what I am actually conceiving is that some substance that is a colourless, odourless, potable liquid—even called 'water'—is not $H_2O$.[35] Of course, I can imagine that too; but that is not what I am imagining when I imagine that water is not $H_2O$: I am imagining something about water. The imagination is *de re*. In the same way, when I imagine that Sarah Palin was the US Vice President after the 2012 US election, I am imagining something about *Palin*. When I imagine that Routley found a box that was empty and not empty, it is *him* that I imagine. And when I imagine that 361 is a prime number (it isn't) I am imagining something about that very number.

*Two*. It might be suggested that this is not the notion of conceivability operative in Hume's dictum, since one who imagines impossibilities is not *clearly* conceiving. If one takes it that one can clearly conceive only what is logically possible, this turns Hume's dictum into an empty tautology—and a useless one, since we may not know what is impossible in this sense. If one is using the word in a more common-sense way, it is something of an insult to say that a logician or mathematician who conceives of impossibilities is not conceiving these things clearly, since it is tantamount to an accusation of confusion. Perhaps, there is some *other* notion of conceivability that satisfies Hume's dictum, and which can serve as a test for possibility. If so, I leave it to others to articulate it. I know of no satisfactory such articulation.[36]

[33] Semantics for a logic of imagination can be found in Niiniluoto 1987, Costa-Leite 2010, and Wansing 201+. These are all variations on *possible*-world semantics, and hence do not allow for imagining the impossible. Even worse, they all require imagination to satisfy certain logical closure conditions. Thus, they all validate the principle that if *A* is imagined, and *A* is logically equivalent to *B*, then *B* is imagined. This is clearly incorrect. *A* is logically equivalent to $(A \wedge C) \vee A$, but I can imagine that Sherlock Holmes lived in Baker St without imagining that (Sherlock Holmes lived in Baker St and $E = mc^2$, or Holmes lived in Baker St). Nothing about Special Relativity need have crossed my mind at all. It is precisely this to which the Secondary Directive caters. Berto (2012: Ch. 7) has a semantics for conception/representation which uses impossible worlds. He does not require that everything be conceivable, but the semantics does allow for that possibility.

[34] One might also doubt that a person understands indefinitely long sentences of such a language. By the same token, one might doubt that such sentences are really grammatical. One might therefore be inclinded to put the same bounds of finitude on both both.

[35] See Berto 2012: 6.3.2.1 for references and discussion.

[36] Chalmers 2002 constructs an eightfold taxonomy of notions of conceivability, and argues that at least one of these entails possibility: ideal primary positive conceivability. This may well be different from the notion of conceivability I am discussing here—though the circularity in his glosses of these notions make me less than certain. But in any case, one thing is clear: the ideality involved is that of some infinite and infallible a

*Three*. It might be suggested that I am confusing imagination with supposition. One can suppose anything; this does not mean that one can imagine anything. One can indeed suppose anything, but I am not talking about supposition. To suppose something is to assume it, usually for the purpose of drawing conclusions.[37] Imagining does not require this. I ask you to imagine that George Bush likes dressing up in a tutu. I am not asking you to suppose anything, or infer anything—merely to use your imagination. Indeed, we will next turn to imagining objects. These are not even the *kind* of thing that can be supposed.

So let us turn to the intentional predicate. For an agent to conceive of an object (*de re*) is simply for them to bring before the mind a term, $t$, which refers to it.[38] And just as I can conceive of any state of affairs I can describe, I can conceive of any object I can describe, even if it is an impossible one. That is, anything of the form $\langle\varphi\rangle mCt$ is true (at the actual world)—where $m$ is me, $t$ is any name or description, and $xCy$ is '$x$ conceives of $y$'.

It is natural to ask, at this point, what the difference is between a possible object and an impossible one—or better which conditions characterise possible objects and which conditions characterise impossible ones.

Let $A$ be any condition with one free variable, $x$. Then a condition is possible if there is a possible world, $P$, where something satisfies $A$ in $P$. Otherwise it is impossible. (By the Primary Directive, any condition is satisfied at some worlds.) Thus, if one takes classical logic to be correct, and $A$ is an inconsistent condition, then it will be an impossible one, and $\varepsilon xA$ will be an impossible object.[39]

Finally, how are matters affected if the Primary Directive is satisfied in possible worlds (for example, if the correct logic is *FDE*)? Then every state of affairs is logically possible. By the same token, every condition is realised in a possible world, so there are no impossible conditions. So whatever $A$ is, $\varepsilon xA$ is a possible object. If this is the case, there is a certain irony here. One must agree with the quote from Hume! If everything is logically possible, then anything 'the mind clearly conceives' is logically possible! Of course, that is not what he meant. Perhaps there is a lesson here.[40]

## 5. Conclusion

The Norwegian explorer Fridtjof Nansen said: "The difficult is what takes a little time; the impossible is what takes a little longer".[41] Philosophy plays the long game. The impossible has always been a marginalised character in Western philosophy. The infinite had always been a marginalised character in mathematics until the time of Cantor. But just as Cantor provided an understanding of the mathematical structure of the infinite, modern logic—especially paraconsistent logic—has provided an understanding of the mathematical structure of the impossible. One can hardly pretend that this is an achievement on the scale of

priori reasoner—not a very useful notion for mere mortals.

[37] *OED*, *to suppose*: 'to think or assume that something is true or probable but lack proof or certain knowledge', 'used to introduce a hypothesis and imagine its development'.

[38] See Priest 1995: 4.8. Again, I am assuming that the agent understands the term '$t$'.

[39] Here, $\varepsilon$ is the indefinite description operator: a (particular) object such that.$\Diamond$

[40] As the old saying goes: be careful of what you wish for; you might just get it.

[41] Cohen and Cohen 1992: 291.

Cantor's (at least so far). However, I think that it has the potential to open people's eyes in the same way. Maybe even wider.[42]

## References

Berto, F. 2012, *Existence as a Real Property*, Dordrecht: Springer.

Berto, F. 2013, "Impossible Worlds", in Zalta, E. (ed.), *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/impossible-worlds. ◇

Bueno-Soler, J. 2012, "Models for Anodic and Cathodic Multimodalities", *Logic Journal of the IGPL*, 20, 458-76.

Carnielli, W., Coniglio, M. and Marcos, J. 2007, "Logics of Formal Inconsistency", in Gabbay, D. and Guenthner, F. (eds.), *Handbook of Philosohical Logic*, 2nd ed., Dordrecht: Kluwer, 1-93. ◇

Carnielli, W. and Pizzi, C. 2008, *Modalities and Multimodalities*, Berlin: Springer.

Chalmers, D. 2002, "Does Conceivability Entail Possibility?", in Szabó, T. and Hawthorne, J. (eds.), *Conceivability and Possibility*, Oxford: Oxford University Press, Ch. 3.

Cohen, J.M. and M.J. 1992, *The New Penguin Dictionary of Quotations*, London: Penguin.

Costa-Leite, A. 2010, "Logical Properties of Imagination", *Abstracta*, 6: 103-16.

Garson, J. 2014, "Modal Logic", in Zalta, E. (ed.), *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/logic-modal.

Hume, D. 1739-40, *A Treatise of Human Nature*, Selby-Bigge, L. and Nidditch, P. (eds.), 2nd ed., Oxford: Clarendon Press, 1978.

Knuuttila, S. 2013, "Medieval Theories of Modality", in Zalta, E. (ed.), *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/modality-medieval.

Lightner, D.T. 1997, "Hume on Conceivability and Inconceivability", *Hume Studies*, 23, 113-32.

Menzel, P. 2013, "Possible Worlds", in Zalta, E. (ed.), *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/possible-worlds.

Mortensen, C. 1989, "Anything is Possible", *Erkenntnis*, 30, 319-37.

Niiniluoto, I. 1985, "Imagination and Fiction", *Journal of Semantics*, 4, 209-22.

Priest, G. 1987, *In Contradiction*, Dordrecht: Martius Nijhoff; 2nd ed., Oxford: Oxford University Press, 2006.

Priest, G. 1995, *Beyond the Limits of Thought*, Cambridge: Cambridge University Press; 2nd ed., Oxford: Oxford University Press, 2002.

Priest, G. (ed.) 1997a, *Notre Dame Journal of Formal Logic*, 38 (4), Special Issue on impossible worlds.

[42] Versions of this paper were given at the Fordham University Metaphysics and Mind Group, the Departments of Philosophy at the University of Amsterdam and the Australian National University, and the conference *Thinking the Impossible*, at the University of Turin. Thanks go to those present for their helpful comments. Thanks, too, go to Hartry Field for comments on an earlier draft.

This paper has already appeared on *Philosophical Studies*, 173 (10), 2016, 2649-62: many thanks to the editors of *Philosophical Studies* for having agreed to republish the paper in this journal.

Priest, G. 1997b, "Sylvan's Box", in Priest (1997a), 573-82; repr. in Priest 2005, Ch. 6.6.

Priest, G. 2005, *Towards Non-Being: The Logic and Metaphysics of Intentionality*, Oxford: Oxford University Press.

Priest, G. 2006, *Doubt Truth to Be a Liar*, Oxford: Oxford University Press.

Priest, G. 2008, *Introduction to Non-Classical Logic: From If to Is*, 2nd ed., Cambridge: Cambridge University Press.

Smith, R. 2011, "Aristotle's Logic", in Zalta, E. (ed.), *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/aristotle-logic.

Wansing, H. 201+, "Remarks on the Logic of Imagination", *Synthese*, to appear.

Yablo, S. 1993, "Is Conceivability a Guide to Possibility?", *Philosophy and Phenomenological Research*, 53, 1-42.

# Counterpossibles in Semantics and Metaphysics

*Timothy Williamson*

*University of Oxford*

*Abstract*

This paper defends from recent objections and misunderstandings the orthodox view that subjunctive conditionals with impossible antecedents are true. It explains apparent counterexamples as cases where a normally reliable suppositional heuristic for assessing conditionals gives incorrect results, which some theorists take at face value.

*Keywords*: conditionals, counterfactuals, counterpossibles, supposition, heuristics.

If whales were fish, their behaviour would differ from what it actually is. If whales were fish, their behaviour would be just as it actually is. Those sound like genuine alternatives. Yet, since whales are by nature mammals, they presumably *could not* have been fish; that would be contrary to their nature. Thus both conditionals are counterpossibles, counterfactual conditionals with impossible antecedents. Semantic orthodoxy makes all counterpossibles true. So the two conditionals were true, and not mutually exclusive after all. Is orthodoxy about counterpossibles correct? The problem is not just how best to tidy up an unimportant little corner of the logic and semantics of counterfactuals. It has significant theoretical and methodological ramifications in several directions. This paper defends orthodoxy against recent objections, and explains recalcitrantly unorthodox appearances by our pre-reflective reliance on a fallible heuristic in assessing conditionals.

## 1. What is at Stake

A counterfactual is a conditional sentence like the so-called subjunctive 'If this were so, that would be so' and unlike the indicative 'If this is so, that is so'. Typically, we use counterfactuals to talk about what would have happened if something had been different from how it actually was. Still, despite the etymology, a counterfactual may have a true antecedent; 'If she were depressed, that would explain her silence' does not imply that she is not depressed. But a counterpossible is a counterfactual whose antecedent is impossible, so false.

*Argumenta* 2, 2 (2017): 195-226
ISSN 2465-2334


DOI 10.23811/42.arg2017.wil

What kind of impossibility is relevant? It is not epistemic. For consider this counterfactual:

(1) If thinking had never occurred, science would have flourished.

The antecedent of (1) is epistemically impossible, because it is incompatible with something we know: that we think. But that does not make the antecedent of (1) impossible in the relevant sense. Presumably, the universe could have been lifeless and thoughtless forever. In that case, science would *not* have flourished. Defenders of orthodoxy should agree that (1) is false. The special theoretical problem in evaluating 'If this had been so, that would have been so' arises when this *could not have been* so. The relevant sort of possibility is objective rather than epistemic or subjective. Moreover, what matters is the most inclusive sort of objective possibility, which we may call *metaphysical possibility*. For the special theoretical problem in evaluating a counterfactual does not arise when, although the antecedent could not *easily* have been so, it could still have been so.[1] Fortunately, the main issues about counterpossibles are not too sensitive to the precise shade of modality at issue.

Some subjunctive conditionals arguably have an epistemic reading (Edgington 2008, Vetter 2016). This paper is mainly concerned with non-epistemic readings of subjunctive conditionals, though some of its conclusions may generalize to epistemic readings too. It is sometimes hard to tell whether one is reading a conditional epistemically or not; the reader may wish to bear that issue in mind.

One view is that counterpossibles suffer some sort of presupposition failure (von Fintel 1998). If the presupposition were semantic, they might even lack a truth-value. However, that is too drastic. In explaining how one can tell that whales are not fish, someone might say 'If whales were fish, they would behave quite differently'. That utterance is felicitous even if it is common ground in the conversation that whales could not really have been fish. Of course, once it is common ground that the antecedent of a counterfactual is impossible, the conditional will often lose any conversational relevance it might previously have had, and so be infelicitous to utter. Thus there may be a defeasible pragmatic presupposition that the antecedent is possible, but it would not enter the semantics.

In discussing the semantics of the subjunctive conditional, I will treat it as a sentential operator, making a complex sentence $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ out of simpler sentences $\alpha$ and $\beta$. That may well be an over-simplification. On Kratzer's influential account (Kratzer 2012), 'if' restricts other sentential operators rather than being one in its own right. The arguments of this paper can be transposed to that more sophisticated setting.

It is convenient, though not crucial, to put the problem of counterpossibles in terms of possible worlds. We read 'possible' as *metaphysically possible*, the relevant standard for present purposes. The orthodox evaluation of the counterfactual $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ depends on the truth-value of $\beta$ at relevant possible worlds at which $\alpha$ is true. But what happens if $\alpha$ is true at *no* possible worlds?

[1] For more on the conception of metaphysical possibility as the most inclusive sort of objective possibility see Williamson 2016. One consequence discussed there is that, where $\Diamond$ expresses metaphysical possibility, the S4 principle $\Diamond\Diamond\alpha \supset \Diamond\alpha$ holds (contrary to Salmon 1989), for if $\Diamond$ expresses a sort of objective possibility, so does $\Diamond\Diamond$. Correspondingly, $\Box\Box\alpha$ is no stronger than $\Box\alpha$ for this modality.

The classic semantic accounts of counterfactuals, going back to Stalnaker (1968) and Lewis (1973), use the framework of compositional intensional semantics, and more specifically of possible worlds semantics (for simplicity in what follows, we hold fixed parameters such as the context and the assignment of values to variables, and leave them tacit). In this setting, we can explain why it is so natural to make all counterpossibles true. It is no optional peculiarity of Stalnaker's semantics, or Lewis's.

We equate the *intension* $|\alpha|$ of a sentence $\alpha$ with the set of possible worlds at which $\alpha$ is true. By hypothesis, the intension of a counterfactual is a function of the intensions of its antecedent and consequent:

(2) $|\alpha \mathrel{\Box\!\!\rightarrow} \beta| = f(|\alpha|, |\beta|)$

We can be more specific, for all that should matter about the consequent is at which possible worlds *where the antecedent is true* the consequent is also true, in other words, the intersection of the intension of the antecedent with the intension of the consequent. That is just the restricting effect of the antecedent. On that assumption, (2) implies (3):

(3) $|\alpha \mathrel{\Box\!\!\rightarrow} \beta| = f(|\alpha|, |\alpha| \cap |\beta|)$

The truth-value of the consequent at worlds where the antecedent is false should be irrelevant to the truth-value of the conditional, for it concerns only what would hold if its antecedent held. But if the antecedent is true at no possible worlds, then the consequent is true at no possible worlds at which the antecedent is true; (3) yields (4):

(4) If $|\alpha| = \{\}$ then $|\alpha \mathrel{\Box\!\!\rightarrow} \beta| = f(\{\}, \{\})$

Given (4), all counterpossibles have the same intension: they are indiscriminate. That does not yet decide between making them all true and making them all false. However, we want any counterfactual (counterpossible or not) whose consequent merely repeats its antecedent to be a trivial necessary truth, true throughout the set of all possible worlds W:

(5) $|\alpha \mathrel{\Box\!\!\rightarrow} \alpha| = W$

Together, (4) (with $\beta = \alpha$) and (5) require $f(\{\}, \{\})$ to be W. Putting that back into (4), we get:

(6) If $|\alpha| = \{\}$ then $|\alpha \mathrel{\Box\!\!\rightarrow} \beta| = W$

In other words, all counterpossibles are necessary truths, and so truths. This is just the conclusion reached by Stalnaker, Lewis, and their successors in the mainstream of intensional semantics.[2]

Similar arguments can be made in the modal object-language, without reference to worlds. For example, instead of (2) we can just require that counterfactuals with necessarily equivalent antecedents and necessarily equivalent consequents are themselves necessarily equivalent:

(7) $(\Box(\alpha \equiv \alpha^*) \wedge \Box(\beta \equiv \beta^*)) \supset \Box((\alpha \mathrel{\Box\!\!\rightarrow} \beta) \equiv (\alpha^* \mathrel{\Box\!\!\rightarrow} \beta^*))$

[2] For Stalnaker (1968) there is a small wrinkle. For convenience, he postulates an impossible 'absurd world' $\lambda$ where everything is true, and equates the truth-value of a counterpossible with the truth-value of its consequent at $\lambda$ (the only world where its antecedent is true). However, the effect is the same, because the consequent is automatically true at $\lambda$. His semantics is still compositional with respect to intensions defined over possible worlds, because $\lambda$ does not discriminate between sentences with the same intension.

A plausible auxiliary assumption to complete the argument is simply that conjunctions necessarily counterfactually imply their conjuncts (if this and that were so, this would be so):

(8) $\Box((\alpha \wedge \beta) \mathrel{\Box\!\!\rightarrow} \alpha)$

Together, (7) and (8) imply (9), the analogue in the object-language of (6):[3]

(9) $\Box\neg\alpha \supset \Box(\alpha \mathrel{\Box\!\!\rightarrow} \beta)$

A much simpler argument in the object-language for the truth of counterpossibles just uses a plausible and attractive assumption linking metaphysical modality to counterfactuals. It is that strict implication materially implies counterfactual implication:

(10) $\Box(\alpha \supset \beta) \supset (\alpha \mathrel{\Box\!\!\rightarrow} \beta)$

If α *could not* hold without β, α *would not* hold without β. Then, by elementary modal logic, an impossibility strictly implies anything:

(11) $\Box\neg\alpha \supset \Box(\alpha \supset \beta)$

By transitivity, (10) and (11) entail (12):

(12) $\Box\neg\alpha \supset (\alpha \mathrel{\Box\!\!\rightarrow} \beta)$

If one assumes the necessitation of (10), one can also derive the necessitation of (12).

Elsewhere, I have used (10) in deriving equivalents of metaphysical modalities in counterfactual terms, as part of an argument for understanding our cognitive capacities for handling metaphysical modalities as a by-product of our cognitive capacities for handling counterfactual conditionals (Williamson 2007: 156-77). This is one of several ways in which issues about counterpossibles have significant knock-on effects for more general philosophical questions.

Thus strong theoretical pressures push towards orthodoxy about counterpossibles. It is required by the standard simple and natural approach to the semantics of counterfactuals, and it contributes to a simple and natural picture of how counterfactuals and metaphysical modality fit together. Nevertheless, those pressures are not obviously irresistible. If one is willing to countenance impossible worlds in addition to possible worlds, one might be able to retain the world-based semantic framework for counterfactuals while rejecting orthodoxy about counterpossibles (Nolan 1997, Brogaard and Salerno 2013, Kment 2014). For if the semantic value of a counterfactual is sensitive to its behaviour at impossible worlds, that makes it easier to deny assumptions such as (2), (7), and (10).[4] Alternatively, one might seek a different semantic framework for counterfactuals that is less congenial to orthodoxy about counterpossibles. The issues must be examined in more depth.

[3] Proof: Substitute $\beta \wedge \neg\beta$ for $\alpha^*$ and β for $\beta^*$ in (7). By elementary propositional modal reasoning (which can be carried out in the weakest normal system K), the result reduces to $\Box\neg\alpha \supset \Box((\alpha \mathrel{\Box\!\!\rightarrow} \beta) \equiv ((\beta \wedge \neg\beta) \mathrel{\Box\!\!\rightarrow} \beta))$. Substitute β for α and $\neg\beta$ for β in (8). The result is $\Box((\beta \wedge \neg\beta) \mathrel{\Box\!\!\rightarrow} \beta)$. Further elementary modal reasoning then yields (9).

[4] If metaphysical possibility obeys the S4 axiom, as suggested in n. 1, then in the metaphysical sense impossible worlds are not even possibly possible, or possibly possibly possible, or …; their impossibility is of a more radical sort.

The focus of resistance to orthodoxy about counterpossibles is usually on alleged counterexamples. Here is one (based on Nolan 1997). What are the truth-values of (13) and (14)?

(13) If Hobbes had (secretly) squared the circle, sick children in the mountains of South America at the time would have cared.

(14) If Hobbes had (secretly) squared the circle, sick children in the mountains of South America at the time would not have cared.

If one responds in a theoretically unreflective way, the natural snap answers are that (13) is false and (14) true. The sick children in the mountains of South America were in no position to know about Hobbes's secret geometrical reasoning thousands of miles away; even if they had known, they had far more urgent things to care about. But, as we now know, the circle cannot be squared. The shared antecedent of (13) and (14) is metaphysically impossible. So orthodoxy makes (13) and (14) alike true. Thus, critics allege, (13) is a counterexample to orthodoxy: a false counterpossible. Such examples can be multiplied without limit.

The temptation to deny (13) and many similar counterpossibles is strong. But such inclinations are not always veridical. For the time being, we may treat them as defeasible evidence against orthodoxy about counterpossibles. Later, we will assess the prospects for an error theory about those inclinations.

For some metaphysicians, rejecting orthodoxy also has more theoretical attractions. Here is an example. Nominalists crave the scientific advantages that platonists gain from quantifying over numbers and other abstract objects. How to emulate them? A common strategy, in this and similar cases, is *fictionalist*. One treats the envied rival metaphysical theory as a useful fiction. The proposal deserves to be taken seriously only if accompanied by a properly worked-out account of how reasoning on the basis of a fiction can nevertheless be a reliably truth-preserving way of getting from non-fictional premises to a non-fictional conclusion. For instance, if one reasons validly from true premises purely about concrete reality plus a false (by nominalist lights) auxiliary mathematical theory about abstract objects to a conclusion purely about concrete reality, the conclusion needs to be true too (Field 1980). But why should it be true? One way of implementing the fictionalist strategy is to use counterfactuals. The nominalist reasons in effect about *how things would be if the mathematical theory were to obtain and concrete reality were just as it actually is*. Thus the conclusion corresponds to this counterfactual:

(15) $(M \wedge A) \mathrel{\square\!\!\rightarrow} C$

Here $M$ is the platonist mathematical theory, $A$ says that concrete reality is just as it actually is, and $C$ says something purely about concrete reality. Thus, the truth of the counterfactual seems to guarantee the truth of its consequent, even though its antecedent is false (by nominalist lights), because the relevant counterfactual worlds are the same as the actual world with respect to concrete reality, which $C$ is purely about. The trouble is that the nominalist may well regard platonism as not just *false* but *metaphysically impossible*: for instance, the structure of the hierarchy of pure sets (if any) seems to be a metaphysically non-contingent matter.[5] For such a nominalist, $M$ is impossible, so the counterfactual

[5] Field (1989) treats Platonist mathematics as contingently false, but with respect to a logical rather than metaphysical modality. Such a logical modality arguably falls outside

(15) is a counterpossible. But, given orthodoxy about counterpossibles, the impossibility of the antecedent guarantees the truth of the counterpossible, irrespective of its consequent, so the truth of (15) is insufficient for the truth of *C*. Fictionalists who implement their strategy with counterfactuals and regard the rival metaphysical theory as a useful but impossible fiction have therefore been compelled to deny orthodoxy about counterpossibles (for instance, Dorr 2008).

We can articulate the relation between the move from (15) to *C* and orthodoxy about counterpossibles. The natural route from (15) to *C* is this. Suppose that *C* is (actually) false. Since *A* says that concrete reality is just as it actually is and *C* says something purely about concrete reality, $\neg C$ does too. Hence $\neg C$ is in effect part of what *A* says. Thus the opposite counterfactual holds:

(16) $(M \wedge A) \mathrel{\Box\!\!\rightarrow} \neg C$

If (16) entails the negation of (15) we can therefore derive $\neg$(15) from $\neg C$, and so *C* from (15) by contraposition. Conversely, we can derive (15) from *C* just as we derived (16) from $\neg C$ (without relying on the mutual exclusion of counterfactuals). But orthodoxy rejects the assumption that (15) and (16) exclude each other, for both are true if their shared antecedent is impossible.

Most orthodox theorists hold that opposite counterfactuals such as (15) and (16) are both true *only* if they are counterpossibles.[6] For they accept the following two principles. First, counterfactuals distribute over conjunction in the consequent:

(17) $(\alpha \mathrel{\Box\!\!\rightarrow} (\beta \wedge \gamma)) \equiv ((\alpha \mathrel{\Box\!\!\rightarrow} \beta) \wedge (\alpha \mathrel{\Box\!\!\rightarrow} \gamma))$

This holds on any standard semantics for counterfactuals. Second, no metaphysical possibility counterfactually implies a metaphysical impossibility:[7]

(18) $(\alpha \mathrel{\Box\!\!\rightarrow} \beta) \supset (\Diamond\alpha \supset \Diamond\beta)$

If something is impossible which would obtain if something else were to obtain, then the latter is impossible too. From (17) and (18) we can easily derive that the conjunction of opposite counterfactuals implies the impossibility of their antecedent:[8]

(19) $((\alpha \mathrel{\Box\!\!\rightarrow} \beta) \wedge (\alpha \mathrel{\Box\!\!\rightarrow} \neg\beta)) \supset \neg\Diamond\alpha$

Even opponents of orthodoxy about counterpossibles may grant (19), since their unorthodoxy may be confined to counterpossibles, and a counterexample to (19) would require a possible antecedent. What opponents of orthodoxy reject, and proponents accept, is the converse of (19).

the range of objective modalities indicated above. For example, 'Socrates = Plato' is presumably possible in the alleged logical sense but is not metaphysically possible; the necessity of distinctness follows from the necessity of identity in S5, the best candidate for the propositional modal logic of metaphysical modality. See Williamson (2016) for more discussion.

[6] For instance, it is axiom (a4) of Stalnaker (1968).

[7] See Williamson (2007: 156). In the setting of a quasi-Lewisian approach to the semantics of counterfactuals, (18) corresponds to the Strangeness of Impossibility Condition Daniel Nolan is "tempted to impose": "any possible world is more similar (nearer) to the actual world than any impossible world" (Nolan 1997: 550; see 566-67 for more discussion).

[8] Proof: Substitute $\neg\beta$ for $\gamma$ in the right-to-left direction of (17) and $\beta \wedge \neg\beta$ for $\beta$ in (18) and use $\neg\Diamond(\beta \wedge \neg\beta)$.

The apparently minor issue of counterpossibles is thus interrelated with various more central questions in both semantics and metaphysics.[9] However, the question should be separated from some other issues with which it is surprisingly often confused. That is the business of the next section.

## 2. Misconceptions about Orthodoxy

In a recent critique of orthodoxy, Berit Brogaard and Joe Salerno characterize their target thus:

> Counterpossibles are trivial on the standard account. By 'trivial', we mean *vacuously true and semantically uninformative*. Counterpossibles are *vacuously true* in that they are always true; an impossibility counterfactually implies anything you like. And relatedly, they are *uninformative* in the sense that the consequent of a counterpossible makes no contribution to the truth-value, meaning or our understanding of the whole (Brogaard and Salerno 2013: 642).

Of this conjunction, orthodoxy as characterized above corresponds only to the first conjunct, the claim of vacuous truth. Brogaard and Salerno handle even that conjunct somewhat oddly. For instance, they say of the counterpossible (14) above: "The intuition is that [(14)] is true, but non-vacuously" (Brogaard and Salerno 2013: 642-43). By their own definition, the non-vacuous truth of (14) consists only in its truth, which both sides acknowledge, and the falsity of at

[9] But some applications of unorthodoxy about counterpossibles to metaphysics fail even on unorthodox terms. A case in point is Brogaard and Salerno's attempt to revive a modal analysis of essence (using counterfactuals) against well-known objections by Kit Fine (1994). As they explain Fine's argument, "While Kripke's wooden table, Tabby, is necessarily a member of the set {Tabby}, it is not essential to Tabby that it be a member of that set" (Brogaard and Salerno 2013: 646). On their account of the ordinary use of 'essential', "There being *F*s is essential to *x* iff if there were no *F*s then *x* would not exist"; this is intended to allow that there being doctors may be essential to someone whose life they saved since if there were no doctors she wouldn't exist. Their account of the 'philosophical use' is the one relevant to Fine's challenge; on it "There being *F*s is essential to *x* (or *x* is essentially *F*) iff (i) if there were no *F*s then *x* would not exist, and (ii) it is metaphysically necessary that if *x* exists then *x* is *F*" (Brogaard and Salerno 2013: 647). But this gets the Tabby example wrong on their own terms. For substitute 'Tabby' for '*x*' and 'member of a singleton of Tabby' for '*F*'. Then (ii) holds because it is metaphysically necessary that if Tabby exists then Tabby is a member of a singleton of Tabby. As for (i), their discussion assumes a contingentist view of existence, otherwise their account of the ordinary use would fail by their own lights in the doctors case. Hence we may assume that Tabby could have failed to exist. In those circumstances, there would have been no members of a singleton of Tabby, because there would have been no singleton of Tabby. Thus the relevant instance of (i) is a non-counterpossible, because its antecedent is possible; thus (i) behaves normally on their view. Moreover, it is metaphysically necessary that if Tabby exists then there are members of a singleton of Tabby, since Tabby itself is one. Hence (i) is true too. Thus, by Brogaard and Salerno's analysis, in the philosophical sense there being members of a singleton of Tabby is essential to Tabby (or Tabby is essentially a member of a singleton of Tabby), exactly the result they needed to avoid. In response to a different problem, they suggest requiring that "F is essential to *a* only if: there might still be Fs even if *a* hadn't existed" (*Ibid.* n. 9), which handles the counterexample above at the unwise cost of making being Tabby inessential to Tabby.

least one other counterpossible, such as (13).[10] Thus the relevant 'intuition' is not directed at (14) at all, but at some other counterpossible. However, that is a minor point compared to their inclusion of the second conjunct, semantic uninformativeness. For we need to be quite clear that semantic uninformativeness is no part whatsoever of the standard account. Consequently, that counterpossibles are trivial in Brogaard and Salerno's sense is no part whatsoever of the standard account.

To see this, recall that on the standard view of counterfactuals they have, as usual for complex sentences, a compositional semantics. Their meanings are built up out of the meanings of their constituents. For authors such as Stalnaker (1968) and Lewis (1973), the meaning of the counterfactual sentence $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ is built up out of the meanings of the sentences $\alpha$ and $\beta$ combined with the meaning of the counterfactual operator $\Box\!\!\rightarrow$. On a fine-grained conception of meaning, any difference in meaning between the sentences $\beta$ and $\gamma$ makes a difference in meaning between the counterfactuals $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ and $\alpha \mathrel{\Box\!\!\rightarrow} \gamma$, whatever the meaning of $\alpha$.[11] That applies just as much when $\alpha$ is impossible as when $\alpha$ is possible. For instance, the counterpossibles (13) and (14) differ by a 'not' in the consequent, which *ipso facto* makes a difference in meaning between (13) and (14). Thus it is just false that, on the standard view, "the consequent of a counterpossible makes no contribution to the […] meaning […] of the whole".

Equally objectionable is Brogaard and Salerno's claim that, on the standard view, 'the consequent of a counterpossible makes no contribution to […] our understanding of the whole'. For instance, consider these two counterfactuals:

(20) If Plato had been identical with Socrates, Plato would have been snub-nosed.

(21) If Plato had been identical with Socrates, 2 + 2 would have been 5.

We may assume that, by the necessity of distinctness, since Plato and Socrates are distinct, it is metaphysically impossible for them to have been identical. Thus both (20) and (21) are counterpossibles. Nevertheless, we understand them by understanding their constituents and how they are put together. For instance, a failure to understand the constituent 'snub-nosed' prevents one from fully understanding (20), but does not prevent one from understanding (21). Thus, on the standard view, our understanding of the consequent contributes to our understanding of the whole counterpossible. Of course, if one happens to *know* that it is metaphysically impossible for Plato to have been identical with Socrates, and one accepts orthodoxy about counterpossibles, then one can work out that (20) is true even if one does not understand 'snub-nosed', but that is just an instance of the general point that one can know that a sentence states a truth without knowing what it states. A trustworthy and trusted native speaker of Mandarin might utter a sentence of Mandarin and tell me that it states a truth without telling me what truth it states. In any case, someone can understand (20) and (21) without knowing that it is impossible for Socrates to have been identical with Plato. Having spent too much time reading dodgy webpages, he might suspect that Plato *was* identical with Socrates. Alternatively, he might know that

[10] In this paper I make the simplifying assumption that all counterfactuals are true or false in a given context.

[11] Implementations of such a fine-grained conception of meaning have been familiar in the tradition of formal intensional semantics from the beginning; see Carnap (1947: 56-59), and Lewis (1970, Section V).

Plato was distinct from Socrates, but doubt the necessity of distinctness on faulty metaphysical grounds. In general, one can understand a counterpossible without knowing it to be a counterpossible, and one's understanding of it is relevantly like one's understanding of other counterfactuals. All these points arise naturally within the framework of a compositional approach to semantics, such as standard accounts assume.

Of course, the informativeness of a sentence does not supervene purely on its intension in a standard semantic framework. For '2 + 2 =4' and a sentence expressing Fermat's last theorem have the same intension, the set of all possible worlds, even though the latter is more informative than the former. Informativeness is sensitive to *how* the given intension is expressed. But that point does not depend on one's view of counterpossibles.

Here is another way to verify the informativeness of counterpossibles, on the standard view. Suppose that someone is initially uncertain whether Plato is identical with Socrates. Then someone whom he trusts utters (21). Since he knows that its consequent is impossible (2 + 2 could not have been 5), he deduces (using (18)) that the antecedent is impossible too, and thereby comes to know that Plato could not have been identical with Socrates. He has gained useful information from (21).

What of Brogaard and Salerno's claim that, on the standard account, "the consequent of a counterpossible makes no contribution to the truth-value […] of the whole"? At first sight, it looks more defensible, since truth-value is a more coarse-grained feature than either meaning or understanding. However, their claim about truth-values is unwarranted too. The only basis for making it is that, according to standard views, all counterfactuals with impossible antecedents have the same truth-value, because all are true, irrespective of their consequent. But, equally, according to standard views, all counterfactuals with *necessary consequents* have the same truth-value, because all are true, irrespective of their antecedent:

(22) $\Box\beta \supset (\alpha \mathrel{\Box\!\!\rightarrow} \beta)$

Both principles, (12) and (22), are corollaries of the quite general entailment (10) from any strict implication to the corresponding counterfactual; one can also derive the semantic analogue of (22) from (3) and (5). Thus, if the standard account implies that the consequent of a counterfactual with an impossible antecedent makes no contribution to the truth-value of the whole, by parity the standard account also implies that the antecedent of a counterfactual with a necessary consequent makes no contribution to the truth-value of the whole. But that combination is absurd. For consider a counterfactual such as (23) with an impossible antecedent *and* a necessary consequent:

(23) If 6 were prime, 35 would be composite.

By Brogaard and Salerno's style of reasoning, the standard account implies that *neither* the antecedent *nor* the consequent of (23) makes any contribution to the truth-value of (23). That is absurd because, without its specific antecedent and consequent, all that is left of (23) is the bare form of a counterfactual sentence alone, which by itself certainly does not determine a truth-value. Obviously, standard theories of counterfactuals such as Stalnaker's and Lewis's have no such ridiculous consequence. Thus even Brogaard and Salerno's claim that, on the standard account, the consequent of a counterpossible makes no contribution to the truth-value of the whole is unwarranted. Such examples also tell in a

parallel way against their already rejected claims that, on the standard account, the consequent of a counterpossible makes no contribution to the meaning and our understanding of the whole.

It thus betrays a grave misunderstanding of orthodoxy to suppose that it makes counterpossibles semantically uninformative or cognitively trivial. It simply makes them true. The misunderstanding is the source of many objections to orthodoxy, including many (though not all) of Brogaard and Salerno's.[12] One such style of objection is this. According to orthodoxy, (24) is true, because Fermat's Last Theorem is a necessary truth:

(24) If Fermat's Last Theorem were false, 2 + 2 would be 5.

The critic then points out, correctly, that Andrew Wiles could not have simplified his famous proof by merely invoking (24) and thence deducing Fermat's Last Theorem by *reductio ad absurdum*. This does indeed refute the claim that (24) is uninformative or trivial, for given the latter claim it is harmless to rely on (24) in a proof. But it is hopeless as an argument against the claim that (24) is true, for the mere truth of a claim does not permit one to rely on it in a *proof*. For that, the claim must have some epistemically appropriate property: it must be an axiom, or have been already proved, or follow from previous steps in a way clear to expert mathematicians, or something like that. Since (24) has no such epistemically appropriate property, it offers no simplification of Wiles's proof. Thus the objection fails. More generally, assertibility requires some epistemically appropriate status, such as being known by the asserter, for which truth is insufficient. That point applies just as much to counterpossibles as to sentences of any other kind, and fits well with orthodoxy. Failure to appreciate it presumably comes from the confused idea that orthodoxy makes counterpossibles uninformative or trivial.

Of course, in terms purely of possible worlds, considering a counterpossible involves supposing that a member of the empty set of worlds obtains, which looks like a waste of time. But questions of uninformativeness and triviality are cognitive and computational, sensitive not only to sets of worlds but also to the linguistic guises under which they are presented. They may conceal or reveal the impossibility of the antecedent. Whether the hearer consciously accepts the vacuous truth of counterpossibles also makes a difference to their informativeness. The mere fact that a counterfactual conditional has an impossible antecedent tells us very little about its cognitive or computational status. For that, we need to know the linguistic guise of the antecedent.

A subtler misconception about orthodoxy concerns speakers who know the impossibility of the antecedent. Consider (25):

(25) If Hobbes had squared the circle, he would have become Lord Chancellor.

I know that the antecedent of (25) is impossible; given orthodoxy, I know that (25) is true. Epistemically, I am in a position to assert (25) on those grounds. But if I do so in a discussion of seventeenth century English politics, something is obviously amiss. However, that point does not tell against orthodoxy, for orthodoxy can easily explain what is amiss. Given orthodoxy, I was also in a position to assert the more informative and equally relevant (26) instead:

[12] For example, Brogaard and Salerno's criticism of my 2007 in the passage beginning "Interestingly enough" (Brogaard and Salerno 2013: 650-51) depends on neglect of the epistemic informativeness of counterpossibles.

(26) Hobbes could not have squared the circle.

Of course, (25) mentions political matters while (26) does not, but my grounds for asserting (25) make the mention factitious and misleading. Therefore, I should have asserted (26)—or, better, just kept quiet—instead, on Gricean grounds of conversational cooperation (Grice 1989: 26-28). Since I did not, my hearers may assume that I asserted (25) because I knew of some politically significant connection between squaring the circle and the Lord Chancellorship, and so be misled. If my hearers correctly identify my grounds for asserting (25), they will recognize the irrelevance of my contribution. Orthodoxy has no trouble in dealing with such cases, as Lewis knew (1973: 25).

In order to keep one's grip on the implications of orthodoxy, a salutary comparison is between the vacuous truth of counterpossibles and the vacuous truth of empty universal generalizations. The impossibility of the antecedent corresponds to the emptiness of the subject term. For it is widely agreed that 'Every N Vs' is true if and only if the extension of N is a subset of the extension of V.[13] Thus, as a special case, if the extension of N is empty, it is a subset of the extension of V, whatever V is, so 'Every N Vs' is true. Consequently, since there are no golden mountains, (27)-(29) are true, and since there are no unicorns, (30)-(32) are true:

(27) Every golden mountain is a mountain.
(28) Every golden mountain is in Africa.
(29) Every golden mountain is a valley.
(30) Every unicorn is a mammal.
(31) Every unicorn is gentle.
(32) Every unicorn is a fish.

It would be absurd to claim that, on this standard account of the universal quantifier, the predicates in (27)-(32) ('is a mountain', 'is in Africa', 'is a valley', 'is a mammal', 'is gentle', 'is a fish') make no contribution to the truth-value, meaning or our understanding of (27)-(32) respectively. For universal generalizations have the same overall compositional semantic structure whether the subject term is empty or not, just as counterfactual conditionals have the same overall compositional semantic structure whether the antecedent is impossible or not. For the same reason, it would be absurd to claim that, on the standard account, (27)-(32) are cognitively trivial or semantically uninformative. After all, one can understand (27)-(29) without knowing that there are no golden mountains, and one can understand (30)-(32) without knowing that there are no unicorns.[14]

Just as we cannot simplify the proof of Fermat's Last Theorem by invoking (24), so we cannot simplify its proof by invoking (33):

(33) Every counterexample to Fermat's Last Theorem is a number both identical to 2 + 2 and identical to 5.

Although (33) is true, we cannot rely on it in a proof without having already proved it, any more than we can do so in the case of (24). Again, knowing the truth of (34) because one knows that no Englishman squared the circle does not make it conversationally appropriate to assert (34) in a discussion of seventeenth century English politics:

[13] For simplicity, I assume a fixed context and a set-sized domain.
[14] For simplicity, I ignore the possibility that 'unicorn' suffers some form of reference failure worse than mere emptiness of extension.

(34) Every Englishman who squared the circle became Lord Chancellor.

As the preceding examples suggest, our reactions to counterpossibles are often similar to our reactions to similarly constructed universal quantifications. For instance, Brogaard and Salerno (2013: 645) write of "the intuitive falsehood" of (35):

(35) If my shirt had been red and non-red all over, then it would have been green.

A first, unreflective reaction to (35) is indeed sceptical: if it had been red and non-red all over, why should it have been green rather than any other colour? Compare (35) with (36):

(36) Every shirt that is red and non-red all over is green.

A first, unreflective reaction to (36) is equally sceptical: if a shirt is red and non-red all over, why should it be green rather than any other colour? In the case of universal quantification, we have learnt to override such immediate reactions. On reflection, (36) is true because it has no counterexample: no shirt is red and non-red all over and yet not green, because no shirt is red and non-red all over. Perhaps we should learn to override our immediate reactions to counterpossibles like (35) in a similar way. The last section of the paper discusses these issues in more depth.

## 3. Counterlogicals

David Lewis offered a brief independent argument for the truth of counterlogicals, counterfactuals with logically inconsistent antecedents. In principle, a moderate opponent of orthodoxy might claim that, although there are false counterpossibles, there no false counterlogicals. In practice, however, opponents of orthodoxy are typically less moderate: they claim that some counterlogicals are false. For if one is moved by examples of apparently false counterpossibles, one will probably also be moved by examples of apparently false counterlogicals, such as (37):

(37) If not everything were self-identical, everything would be self-identical.

Similarly, Brogaard and Salerno (2013: 643) classify (38) as false:

(38) If intuitionistic logic were the correct logic, then the law of excluded middle would still be unrestrictedly valid.

Although (38) is a slightly trickier case, because its antecedent is inconsistent with metalogic rather than the logic of a non-metalinguistic object-language, for present purposes the difference does not matter.

The dialectical setting is this. Opponents of orthodoxy about counterpossibles typically grant classical logical principles for standard logical constants such as the truth-functors, quantifiers, identity, and even modal operators (not including the counterfactual conditional itself). They also grant standard structural principles about logical consequence, such as the cut rule and monotonicity. In that sense, classical logic is not at stake, though if one prefers a non-classical logic, one can have a similar debate about counterpossibles in that setting too. For present purposes, we may simply assume that both sides in the debate over counterpossibles accept classical logic.

Lewis compresses his argument into a single sentence:

> [I]t seems that a counterfactual in which the antecedent logically implies the consequent ought always to be true; and one sort of impossible antecedent, a self-contradictory one, logically implies any consequent (Lewis 1973: 24).

Here is one way of unpacking Lewis's argument. For convenience, we represent his 'logically implies' by the logical truth of a material implication rather than the logical validity of an argument from premises to conclusion; nothing of present importance hangs on the choice. Logical truth is expressed by '⊢'. Suppose that α is 'self-contradictory':

(39) $\vdash \neg\alpha$

Then, by classical logic, α logically implies any consequent γ:

(40) $\vdash \alpha \supset \gamma$

Lewis endorses the plausible principle he calls 'Deduction within Conditionals' (1973, p. 132). It says that what is counterfactually implied by a given antecedent is closed under deduction:[15]

(DC) If $\vdash (\wedge_{i \in I} \beta_i) \supset \gamma$ then $\vdash (\wedge_{i \in I} (\alpha \mathrel{\Box\!\!\rightarrow} \beta_i)) \supset (\alpha \mathrel{\Box\!\!\rightarrow} \gamma)$

(DC) fits well with our practice of making deductions in developing a supposition to reach a counterfactual conclusion; it also yields the logical truth of (17) (the counterfactual conditional commutes with conjunction in the consequent). The present argument requires only the special case of (DC) with just one premise:

(DC1) If $\vdash \beta \supset \gamma$ then $\vdash (\alpha \mathrel{\Box\!\!\rightarrow} \beta) \supset (\alpha \mathrel{\Box\!\!\rightarrow} \gamma)$

If we substitute α for β in (DC1) and combine the result with (40), we derive (41):

(41) $\vdash (\alpha \mathrel{\Box\!\!\rightarrow} \alpha) \supset (\alpha \mathrel{\Box\!\!\rightarrow} \gamma)$

The reflexivity principle that everything counterfactually implies itself is unproblematic, and an axiom of Lewis's preferred logic of counterfactuals (Lewis 1973: 32):

(R) $\vdash \alpha \mathrel{\Box\!\!\rightarrow} \alpha$

From (41) and (R) we derive (42):

(42) $\vdash \alpha \mathrel{\Box\!\!\rightarrow} \gamma$

Thus, by plausible principles of Lewis's logic of counterfactuals, we can derive (42), the theorem that a hypothesis counterfactually implies anything we like, from (39), the assumption that the hypothesis is logically inconsistent.

Lewis's argument and his conclusion require one minor qualification. For suppose that the object-language contains a rigidifying 'actually' operator A, where the sentence Aα is true at an arbitrary world in a model if and only if α is true at the designated actual world of the model. Suppose further that logical truth is truth at the actual world of every model.[16] Substitute $\alpha \wedge \neg A\alpha$ for α.

[15] Lewis implicitly requires the premise set I to be finite; without the Limit Assumption the infinitary version of the principle is invalid on his semantics (Lewis 1973: 19-21). Informally, however, the infinitary version is almost as plausible as the finitary one; it is valid on Stalnaker's semantics. Lewis also requires the premise set to be nonempty, but the special case with an empty premise set is equally plausible, and valid on his semantics. It boils down to this: if $\vdash \gamma$ then $\vdash \alpha \mathrel{\Box\!\!\rightarrow} \gamma$ (the conjunction of an empty set of conjuncts is a tautology, since a conjunction is false only if at least one conjunct is false).

[16] In the terminology of Davies and Humberstone (1980), this corresponds to real world validity rather than general validity.

Then (43) holds because α and Aα have the same truth-value at the designated actual world of any model:

(43) $\vdash \neg(\alpha \wedge \neg A\alpha)$

But we do not in general want (44):

(44) $\vdash (\alpha \wedge \neg A\alpha) \mathrel{\Box\!\!\rightarrow} (\alpha \wedge \neg\alpha)$

For example, if α means 'It is raining', then in effect (44) makes it a logical truth that if it had rained but not in this world, a contradiction would have obtained. The supposed logical truth is false if it could have rained but is not actually doing so. Since (R) is harmless, the trouble is with (DC), and more specifically with (DC1). They do not apply to languages with fancy modal operators such as 'actually' (if logical truth is truth at the actual world of every model). For such circumstances yield contingent logical truths, logically guaranteed to express truths but not logically guaranteed to express necessary truths. The standard Rule of Necessitation in modal logic fails under the same circumstances:

(RN) If $\vdash \alpha$ then $\vdash \Box\alpha$

For example, although (43) holds, (45) fails, because it could have rained in counterfactual but not in actual circumstances:

(45) $\vdash \Box\neg(\alpha \wedge \neg A\alpha)$

For these familiar reasons, (DC) and (DC1) must be restricted in the same way as (RN), for languages with fancy modal operators. In more ordinary cases, contingent logical truths are not involved, the restrictions are satisfied and this problem for Lewis's argument does not arise.[17]

The problem above for Lewis arises when logical implication is insufficient for strict implication. Ironically, in criticizing his argument, Brogaard and Salerno conflate logical implication with strict implication. They miss not only that problem but also also the moderate option on which there are false counterpossibles but no false counterlogicals, which arises because logical implication is unnecessary for strict implication. They seem to be under the misapprehension that the opponent of orthodoxy about counterpossibles is bound to reject Lewis's argument about counterlogicals (Brogaard and Salerno 2013: 648-49).

Those problems aside, Brogaard and Salerno's response to Lewis's argument is clearly to reject (DC) and (DC1). However, they reject them *only* for counterpossibles. For they say: "All the typical rules governing counterfactuals are valid, when the antecedent is possible" (Brogaard and Salerno 2013: 657). They treat it as a virtue of their account that it preserves a strong logic of counterfactuals with possible antecedents. However, from an abductive perspective, that puts increased pressure on their reasons for rejecting the standard rules in the first place (at least for simple languages without fancy modal operators such as 'actually'). Those reasons had better be robust enough to justify the sacrifice of a strong and simple theory such as Stalnaker or Lewis's logic of counterfactuals, in favour of one patched up with messy adjustments and repairs. In particular, given that counterfactuals have a compositional semantics, we should be suspicious of the idea that their behaviour is radically different on the rare occasions when the antecedent is impossible. For how are we supposed to have come to be using so oddly contoured a conditional?

[17] See Williamson (2006 and 2007: 295-96) for more discussion.

## 4. Objectivity and Opacity

A useful litmus test for the nature of a semantic construction is this: given co-referential inputs, does it produce co-referential outputs? If yes, it is *transparent*. If no, it is *opaque*. Of course, a proper account would work hard to clarify the relevant conception of reference, but for present purposes we can just use that schematic explanation.

To set up an example, consider 'Hesperus' and 'Phosphorus' as co-referential proper names, and 'Hesperus will explode tonight' and 'Phosphorus will explode tonight' as co-referential sentences. A new expression 'Prob' is introduced. It takes sentences as inputs and outputs complex singular terms for 'probabilities' in some sense. Compare two cases.

Case (i): We are authoritatively told that (46) is sometimes true:

(46) Prob(Hesperus will explode tonight) < Prob(Phosphorus will explode tonight).

Case (ii): We are authoritatively told that (47) is always true:

(47) Prob(Hesperus will explode tonight) = Prob(Phosphorus will explode tonight).

Case (i) implies the opacity of 'Prob'. It goes well with a subjective or epistemic reading on which Prob(α) is the credence or evidential probability of α. An agent may in some relevant sense be less confident that Hesperus will explode tonight than that Phosphorus will explode tonight, or have less evidence for the former than for the latter. But case (i) goes poorly with an objective reading of 'Prob', on which Prob(α) is the physical chance of α. How can it be physically less probable that Hesperus will explode than that Phosphorus will explode tonight, when those are the very same state of affairs? By contrast, case (ii) is consistent with transparency. It goes poorly with a subjective or epistemic reading of 'Prob', for why should agents always be exactly as confident in 'Hesperus will explode tonight' as in 'Phosphorus will explode tonight', or always have exactly as much evidence for the former as for the latter? But case (ii) goes well with an objective reading of 'Prob', where the identity of the states of affairs forces the identity of their chances.

A natural explanation of the difference is that objective probabilities concern only the states of affairs themselves, whereas subjective and epistemic probabilities are also sensitive to differences in how agents represent those states of affairs—in modes of presentations, or guises, or something like that. The explanation may well need to be refined in all sorts of ways, but it looks to be at least roughly on the right track. At any rate, constructions that concern epistemic or subjective matters invite an opaque reading, while constructions that concern objective, non-epistemic and non-subjective matters do not invite an opaque reading. For simplicity, we will treat the opaque readings in the former case as the intended ones.

Normally, counterfactuals are on the objective side of this contrast. For instance, (48) and (49) should have the same truth-value:

(48) If Hesperus were to explode tonight, so would the Moon.

(49) If Phosphorus were to explode tonight, so would the Moon.

For the antecedents of (48) and (49) suppose the very same state of affairs to obtain. Normally, counterfactuals on a non-epistemic reading are transparent. They are about the very objects, properties, relations, and states of affairs their

antecedents and consequents are about, not the ways in which agents represent those objects, properties, relations, and states of affairs.

If non-epistemic counterpossibles are uniform in this respect with other non-epistemic counterfactuals, they too are transparent. Although the objective impossibility of the antecedent might pragmatically trigger an epistemic reading, it is quite implausible that on an unequivocal non-epistemic reading the natural language counterfactual conditional should have been programmed all along to make the radical semantic switch to non-transparent behaviour just for the unusual case of an impossible antecedent. Transparency puts significant limits to the programme of using (presumed) counterpossibles to simulate one metaphysical theory from within another, fictionalist-style. For instance, suppose that on the true metaphysical view T, a mental event $m$ is identical with a physical event $p$. On a rival view T*, $m$ is distinct from $p$; T* is false, indeed impossible (by the necessity of identity). Proponents of T use counterpossibles to characterize how things are from the point of view of T*; for instance:

(50) T* $\Box\!\!\rightarrow m \neq p$

Of course:

(51) T* $\Box\!\!\rightarrow m = m$

But since $m = p$, both in fact and from the point of view of T, transparency yields (52) from (51):

(52) T* $\Box\!\!\rightarrow m = p$

Thus the impossible but internally coherent metaphysical theory T* still has mutually inconsistent counterfactual implications ($m = p$ and $m \neq p$), even though proponents of T* may be strict adherents of classical logic.

Unsurprisingly, Brogaard and Salerno deny that counterfactuals are transparent.[18] The question is whether substitution of co-referential terms makes a difference in truth-value only for counterpossibles, or for other counterfactuals too. Unfortunately, it is hard to extract a stable answer from their discussion.[19] The dilemma is this. First, suppose that substitution of co-referential terms makes a difference in truth-value only for counterpossibles. Then counterfactu-

[18] Brogaard and Salerno seem to think that the opacity of counterfactuals can be derived from their hyperintensionality: "Hyperintensional operators do not permit substitutions of co-referring terms *salva veritate*" (Brogaard and Salerno 2013: 650). That is false. For instance, let $\alpha \# \beta$ be true when $\alpha$ and $\beta$ express the same Russellian structured proposition, and false otherwise. Then # is hyperintensional, because sentences with the same intension may express different Russellian propositions, but # is transparent, because (absent other opaque operators) substitutions in its scope of co-referring terms preserve the Russellian proposition expressed.

[19] They write: "Of course, we need to give a principled account of when counterfactuals create opaque contexts. They create opaque context [sic] when the antecedent or consequent which result [sic] from substituting one term for another does not follow a priori from the original. Since we are likely to use 'Clark Kent' and 'Superman' in such a way as to pick out the same individual, 'Clark Kent has the same parents as I do' is an a priori implication of 'Superman has the same parents as I do'. So, substitution is legitimate" (Brogaard and Salerno 2013: 650, n. 11). As stated, this seems to require 'Clark Kent is Superman' to be a priori. Perhaps they mean a priori implications given the relevant identity as an auxiliary premise. In any case, none of this obviously helps with counterpossibles such as (50)-(52), where no peculiar constructions need appear in antecedent or consequent.

als behave in radically different ways depending on the modal status of their antecedent: transparently, like a non-epistemic operator, if it is possible, opaquely, like an epistemic operator, if it is impossible. That suggests an implausibly hybrid semantics. A more uniform treatment is much to be preferred. Alternatively, suppose that substitution of co-referential terms makes a difference in truth-value for counterfactuals with possible antecedents too. Then we should expect some evidence of that, especially in the form of convincing examples. Brogaard and Salerno provide none. Indeed, we may expect them to prefer the first horn of the dilemma, given their already quoted remark that "All the typical rules governing counterfactuals are valid, when the antecedent is possible", made in the context of the reasoning, which they endorse, from (53) and (54) to (55) (2013, p. 657):

(53) If the rocket had continued on that course, it would have hit Hesperus.
(54) Hesperus = Phosphorus.
(55) If the rocket had continued on that course, it would have hit Phosphorus.

Brogaard and Salerno's general approach to the semantics of counterfactuals is to keep the structure of Lewis's overall account but to add lots of impossible worlds, considered simply as sets of sentences, not required to be deductively closed, in order to handle counterpossibles. Of course, some sets of sentences do contain $m = m$ without containing $m = p$, even though the latter is in fact true; the singleton $\{m = m\}$ is an example. They flesh out their semantic theory by adding an account of relative closeness for impossible worlds:

> For any two impossible worlds $w_1$ and $w_2$, $w_1$ is closer to the base world than $w_2$ iff
> (a) $w_1$ does not contain a greater number of sentences formally inconsistent with the relevant background facts (held fixed in the context) than $w_2$ does.
> And if $w_1$ and $w_2$ contain the same number of sentences formally inconsistent with the relevant background facts (held fixed in the context):
> (b) $w_1$ preserves a greater number of a priori* implications between sentences than $w_1$ does (Brogaard and Salerno 2013: 655).

They explain a priori* implication thus: "For a speaker *s* in a context *c*, P *a priori* implies* Q iff for *s* in *c*, Q is a relevant a priori consequence of P" (*Ibid*.).

We need not examine Brogaard and Salerno's semantics in detail. It is hard to believe that one will get a useful measure of closeness by counting numbers of sentences, especially when $w_1$ and $w_2$ both contain a countable infinity of sentences. We are also not told how to evaluate open formulas at impossible worlds, as required in evaluating quantified counterpossibles. For instance, is the open sentence '*x* is bright' true or false at a world *w* containing the sentence 'Hesperus is bright' but not the sentence 'Phosphorus is bright', under the assignment of Hesperus, which is to say Phosphorus, as the value of the variable '*x*'?

For present purposes, a more interesting feature of Brogaard and Salerno's account is the appearance of the distinctively epistemic notion of a priori* implication in the account of relevant closeness for impossible worlds. They leave in place Lewis's account of the relative closeness of possible worlds, which depends on objective, non-epistemic features of the worlds (given a fixed context). Thus their overall relative closeness relation is patched together from epistemic and non-epistemic pieces. It is hard to avoid the impression that the account is

being gerrymandered just to accommodate the marginal case of counterpossibles (see also the definition of validity in Berto, French, Priest, and Ripley 2016). Such a hybrid approach resembles an account of conditional probabilities on which they are purely objective when the conditioning event has a positive unconditional chance, but go epistemic or subjective when the conditioning event has zero unconditional chance: not an attractive option.

In addition to the general implausibility of such hybrid theories, there is a more specific problem. As opponents of orthodoxy like to emphasize, many counterpossibles do not simply crash when uttered. Speakers and hearers handle them in ways very similar to the ways in which they handle non-counterpossible counterfactuals. But if the semantics of counterfactuals is to be done in terms of the relative closeness of worlds, which turns out to work in radically different ways depending on whether the worlds are possible or impossible, then *shouldn't* speakers and hearers handle counterpossibles quite differently from how they handle other counterfactuals? Lewis's relative closeness for possible worlds is not remotely like Brogaard and Salerno's relative closeness for impossible worlds, so how come the same style of cognitive processing works for both? A more uniform account would be more plausible.

Closely related to the opacity of counterfactuals on the impossible worlds approach is the lack of a non-trivial relation of synonymy appropriately related to the compositional semantics. For suppose that the semantics is compositional with respect to some kind of meaning, m-meaning, in the sense that the m-meaning of a complex expression supervenes on the m-meanings of its atomic constituents and the way they are put together. The corresponding relation of synonymy is just sameness of m-meaning. If such m-synonymy is non-trivial, then it should sometimes hold between distinct atomic expressions. To take a standard example, suppose that 'furze' and 'gorse' are m-synonymous. Since the semantics is compositional for m-meaning, substituting one of those words for the other in a complex expression preserves its m-meaning. Thus (56) and (57) are m-synonymous (all occurrences of 'this' and 'it' being coreferential):

(56) If this were furze but not gorse, it would be furze but not gorse.
(57) If this were furze but not gorse, it would be furze but not furze.

But the semantics with impossible worlds will typically evaluate (56) and (57) differently: (56) will trivially be evaluated as true, whereas (57) will usually be evaluated as false because supposing that furze and gorse are distinct natural kinds requires nothing as extreme as supposing a contradiction, even though the former supposition is impossible too. Since (56) and (57) differ in truth-value, they differ in m-meaning, for the truth-values of sentences (in a context) should supervene on their compositional meanings, here m-meanings. This contradicts the previous result that (56) and (57) are m-synonymous. Since parallel considerations apply to any other supposed case of non-trivial m-synonymy, we must deny the initial assumption that m-synonymy is non-trivial. Another way of thinking about the issue is that the use of all sets of sentences as impossible worlds in the semantics for the counterfactual conditional in effect turns it into a quotational context, which is equally unattractive. The upshot is that the impossible worlds approach provides a compositional semantics for the counterfactual conditional only in a rather feeble sense.

## 5. Counterfactual Reasoning by Reductio ad Absurdum

The prize specimens of useful reasoning from an impossible supposition are arguments by *reductio ad absurdum* in mathematics. When we state them in everyday terms, it is natural to use counterfactual conditionals. Here are examples adapted from Lewis (1973: 25):

(58) If $p$ were the largest prime, $p! + 1$ would be prime.
(59) If $p$ were the largest prime, $p! + 1$ would be composite.

They summarize a slight variation on Euclid's proof that there is no largest prime: (58) holds because if $p$ were the largest prime, $p!$ would be divisible by all primes (since it is divisible by all natural numbers from 1 to $p$), so $p! + 1$ would be divisible by none; (59) holds because $p! + 1$ is larger than $p$, and so would be composite if $p$ were the largest prime. To complete the proof, one can use Lewis's principle (DC) to conjoin the consequents of (58) and (59):

(60) If $p$ were the largest prime, $p! + 1$ would be both prime and composite.

Since the consequent of (60) is a contradiction, one can deny the antecedent, and conclude that there is no largest prime.

Of course, one does not strictly *need* to formulate the proof in terms of counterfactual conditionals. One could use material conditionals instead, for all standard mathematical reasoning can be formalized in purely extensional terms. Nevertheless, it is surely legitimate, indeed natural and appropriate, to use counterfactual conditionals. They nicely convey the role of the antecedent in the reasoning, especially when the hearer had already been told that there is no largest prime, and wanted to know why. At the very least, on a good semantic theory, the counterpossibles (58)-(60) should come out *true*, for they are soundly based on valid mathematical reasoning.

It is far from clear that (58)-(60) come out true on Brogaard and Salerno's theory. The point is not merely that, on their account, some impossible worlds include the sentence '$p$ is the largest prime' while excluding the sentence '$p! + 1$ is prime' or '$p! + 1$ is composite' (or both), for that does not show that they are the closest worlds including '$p$ is the largest prime' to the actual world. It is instructive to note Brogaard and Salerno's reaction to a much more elementary argument by *reductio ad absurdum*:

> [I]f 5 + 7 were 13 then 5 + 6 would be 12 (and so by another eleven steps) 0 would be 1, so if the number of right answers I gave were 0, the number of right answers I gave would be 1 (Williamson 2007: 172).

They analyse that compressed argument as using *reductio ad absurdum* but argue that, as it stands, it is not cogent, because it does nothing to show that the nearest impossible world in which the antecedent is true has the required properties (Brogaard and Salerno 2013: 649-50). Exactly the same objection applies to the classic proof that there is no largest prime, cast in counterfactual terms. After all, what mathematician has ever supplemented a standard proof by *reductio ad absurdum* with considerations about the relative closeness to the actual world of various impossible worlds, considered as sets of sentences that need not be deductively closed? The cost of Brogaard and Salerno's rejection of my toy proof is a commitment to rejecting ordinary mathematical proofs by *reductio ad absurdum* cast in counterfactual terms. Nor is there any reason to expect special contextual

factors somehow to rescue the hard mathematical proofs but not my easy one, or to postulate a different meaning for the counterfactual conditional in mathematics from its meaning in philosophy.

An initial reaction might be that Brogaard and Salerno must have played their cards badly, and that the trouble they have got themselves into reveals nothing about the strength of their original hand. But further reflection suggests that the problem for opponents of orthodoxy goes much deeper than that, and does not depend on the details of their view. For consider any non-obvious impossibility $\alpha$ that can be shown, by more or less elaborate deductive reasoning, to lead to an obvious impossibility $\omega$. The general anti-orthodox strategy is to be charitable by evaluating counterfactuals with $\alpha$ as the antecedent at impossible worlds or situations not closed under such reasoning, precisely in order to falsify counterpossibles such as $\alpha \mathrel{\Box\!\!\rightarrow} \omega$. But those are exactly the counterpossibles one needs to assert in articulating the argument by *reductio ad absurdum* against $\alpha$. Thus the point generalizes, for instance to the use of counterlogical worlds in Nolan (1997).

One fallback is to concede orthodoxy for counterlogicals, countermathematicals, and the like, but still reject it for some countermetaphysicals. But not only is that a long retreat, it risks undermining the motivation for the residual resistance to orthodoxy in the first place, since the motivating features of the examples are already present in the case of counterlogicals. If those features are somehow illusory, who is to say that similar illusions are not also at work with countermetaphysicals? Thus the fallback is unstable.

Another response is to say that formulating mathematical arguments in counterfactual terms was always loose speech. One must then explain why those formulations *seemed* so compelling to the most rigorous and precise reasoners in our community—mathematicians and logicians—in effect, why they mistook loose speech for strict speech. That would be some sort of error theory. Such a response also risks instability, for if anti-orthodox theorists need an error theory, what is to stop orthodox theorists from applying their own error theory to the other side's considerations, and escaping the messy complications of anti-orthodoxy?

Yet another defensive move is to invoke context-dependence, arguing that (60) is true in the context of standard mathematical proof (contrary to Brogaard and Salerno's reaction) but false in some other contexts. This move might concede a reading of counterfactuals on which they can be used to define metaphysical modalities, while also providing another reading of them on which they can be used for fictionalist purposes in metaphysics. Of course, the appeal to context-dependence is a generic way of rejecting the validity of just about any deductive argument. To carry weight, it needs more specific support. It is not controversial that counterfactuals exhibit some degree of context-dependence in which factors are held fixed; if Julius Caesar had been in command during the Korean War, would he have used catapults or nuclear weapons? But that makes it all the more striking how hard it is to hear the mathematical argument for (60) as unsound. If context determined whether standard logic and mathematics were to be held fixed, (60) might be expected to trigger a shift to a context in which they are not, because it brings out so clearly that the antecedent is untenable without such a shift. Yet the argument for (60) remains compelling. Although one might be able to browbeat ordinary hearers into questioning the mathematical arguments, that is very different from the smooth, unreflective

transitions characteristic of ordinary context-dependence. Without adequate motivation, the appeal to context-dependence is just an all-purpose objection to any valid argument. That context can shift the relative ranking of given worlds does not mean that it can shift which worlds are being ranked at all.

Mathematical arguments by *reductio ad absurdum* are amongst the best arguments for counterpossibles we have. They tell us that if something non-obviously impossible were the case, something obviously impossible would be the case. We should accept the conclusions of those mathematical proofs. They provide strong evidence for orthodoxy. But how can we explain away the strongest evidence *against* orthodoxy, all the seemingly clear examples of false counterpossibles?

## 6. An Error Theory of Apparently False Counterpossibles

Processing a non-obvious counterpossible typically *feels* very like processing a non-counterpossible counterfactual. Consider (13), a good example of a seemingly false counterpossible ('If Hobbes had (secretly) squared the circle, sick children in the mountains of South America at the time would have cared'). What goes on when we process it? In my case, before I consciously apply any theoretical considerations, it is something like this. I imagine Hobbes doing geometry in the secrecy of his room. I ask myself whether sick children in the mountains of South America at the time would have cared. I answer in the negative, because there was no way for them to have known about Hobbes's doings at the time, and even if they had known, they would hardly have cared. In the first instance, I assent to (14), the opposite counterfactual to (13), with the same antecedent but the negation of the consequent ('If Hobbes had (secretly) squared the circle, sick children in the mountains of South America at the time would *not* have cared'). My immediate inclination is then to deny (13), as excluded by (14). So far, in this case, the impossibility of the antecedent has played *no role whatsoever*. That is not to deny that I imagine Hobbes (secretly) squaring the circle. In some minimal, vague, unspecific way I do imagine him squaring the circle, but I could imagine him carrying out some genuine geometrical construction in much the same way. Now, in my case, theory kicks in. I remind myself that squaring the circle is impossible, and that opposite counterfactuals may both be true when their shared antecedent is impossible. I therefore countermand my inclination to deny (13).

Schematically, what we seem to do when we assess the counterfactual $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ is this. First we counterfactually suppose $\alpha$. Then, if within the scope of the counterfactual supposition we accept $\beta$, outside the scope of that supposition we accept $\alpha \mathrel{\Box\!\!\rightarrow} \beta$. Similarly, if within the scope of the counterfactual supposition we reject $\beta$, outside the scope of that supposition we reject $\alpha \mathrel{\Box\!\!\rightarrow} \beta$. More generally, whatever assessment we make of $\beta$ within the scope of the counterfactual supposition of $\alpha$ we make of $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ outside the scope of that supposition; call that the *suppositional procedure*. There is extensive psychological evidence that we tend to evaluate conditionals by evaluating their consequents on the supposition of their antecedents, with only subtle differences in treatment between indicatives and consequents.[20]

[20] Evans, Handley, and Over 2005 provides a useful overview of the evidence for a suppositional account of the evaluation of conditionals, though they may sometimes be too

From an orthodox perspective, over-hasty applications of the suppositional procedure to counterpossibles produce false judgments. At first sight, that looks like a golden opportunity for unorthodox theorists to motivate their semantics. If an unorthodox semantic theory vindicates our way of assessing counterfactuals, while orthodox theories do not, that is a strong point in favour of unorthodoxy.

On further reflection, however, any prospect of fully vindicating the suppositional procedure goes dim. The first reason is just the use of counterfactuals in articulating mathematical proofs by reductio ad absurdum, discussed above. In a normal mathematical context, we counterfactually suppose that $p$ is the largest prime, and in the scope of that supposition deduce the absurd conclusion that $p! + 1$ is both prime and composite. Still within the scope of the counterfactual supposition, we reject that as absurd. But we do not reject the corresponding counterfactual conditional ((60) above). Rather, we accept it in explaining our reasoning. Thus, on our best behaviour, we *contravene* the suppositional procedure.

A similar result follows given any sentence $\omega$ so absurd that we must reject it under any supposition whatever. For then we reject $\omega$ under the counterfactual supposition of $\omega$, so the suppositional procedure tells us to reject the corresponding counterfactual conditional $\omega \mathrel{\Box\!\!\rightarrow} \omega$. But on reflection we do not; we accept $\omega \mathrel{\Box\!\!\rightarrow} \omega$ because anything of the form $\alpha \mathrel{\Box\!\!\rightarrow} \alpha$ is a logical truth, even on the usual versions of impossible world semantics. On a variant of this theme, we might take $\omega$ to be 'Everything should actually be rejected'. In the scope of that counterfactual supposition, perhaps one should reject even $\omega$ itself, but outside the scope of that supposition one should not reject $\omega \mathrel{\Box\!\!\rightarrow} \omega$.

The suppositional procedure also gets into trouble with probabilistic assessments. On its most natural application to such cases, the procedure involves equating one's credence in the conditional $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ with one's conditional credence in $\beta$ on $\alpha$:

(Cond) $\mathrm{Prob}(\alpha \mathrel{\Box\!\!\rightarrow} \beta) = \mathrm{Prob}(\beta \mid \alpha)$

There is evidence that, in general, our assessment of the probability of a conditional is highly correlated with our assessment of the conditional probability of its consequent on its antecedent (Evans, Over, and Handley 2005). But (Cond) breaks down for counterlogical $\alpha$. Under the standard equation of the conditional probability $\mathrm{Prob}(\beta \mid \alpha)$ with the ratio $\mathrm{Prob}(\alpha \wedge \beta)/\mathrm{Prob}(\alpha)$ of unconditional probabilities, the conditional probability is undefined when $\mathrm{Prob}(\alpha)$ is zero, as the axioms of probability require it to be when $\alpha$ is a contradiction. If instead we treat conditional probabilities as primitive, we can sometimes assign $\mathrm{Prob}(\beta \mid \alpha)$ a value even when $\mathrm{Prob}(\alpha) = 0$, but we still cannot do so when $\alpha$ holds at *no* point in the probability space, on pain of violating basic principles of conditional probability. For $\mathrm{Prob}(\beta \mid \alpha)$ should be 1 whenever $\alpha$ entails $\beta$ in the sense that $\beta$ holds at every point in the space at which $\alpha$ holds, so for vacuous $\alpha$ $\mathrm{Prob}(\beta \mid \alpha) = \mathrm{Prob}(\neg\beta \mid \alpha) = 1$ for any $\beta$, violating the principle that $\mathrm{Prob}(\neg\beta \mid \alpha) = 1 - \mathrm{Prob}(\beta \mid \alpha)$. Thus the structure of probability theory rules out vacuous conditional probabilities. Rejecting standard principles of conditional probability to allow for vacuous conditional probabilities would be a fool's bargain.

quick in passing between psychological claims and semantic ones. See Evans and Over (2004: 113-31).

In any case, (Cond) faces problems even when α is possible. The equation of probabilities of conditionals is more attractive for indicative conditionals than for subjunctive ones like $\alpha \mathrel{\Box\!\!\rightarrow} \beta$; combining both equations commits one to equating the probability of a subjunctive conditional with the probability of the indicative conditional with the same antecedent and consequent, thereby erasing a significant distinction. Moreover, a long line of impossibility results, initiated by David Lewis, shows that under very general conditions there cannot be conditional propositions whose probabilities always equal the corresponding conditional probabilities (Lewis 1976, Hájek 2011).

Would-be vindicators of the suppositional procedure might respond to these problems by denying that the probability of β under the counterfactual supposition of α is the conditional probability of β on α. They might instead identify it with the probability of β under some other probability distribution $\text{Prob}_\alpha$ determined by α:

(Cond*) $\text{Prob}(\alpha \mathrel{\Box\!\!\rightarrow} \beta) = \text{Prob}_\alpha(\beta)$

Since the probability of α itself under the counterfactual supposition of α should be 1, $\text{Prob}_\alpha(\alpha) = 1$, and (Cond*) yields the same result since $\text{Prob}(\alpha \mathrel{\Box\!\!\rightarrow} \alpha) = 1$. Thus not even (Cond*) solves the problem of counterlogicals, since $\text{Prob}_\alpha$ was supposed to be a probability distribution, which requires it to assign probability 0, not 1, to contradictions. Rejecting standard axioms of the probability calculus for the sake of one's favoured treatment of counterpossibles would be in the worst methodological taste, since the explanatory power of probability theory depends on its mathematical strength, which relies on those axioms.

For possible α, (Cond*) does not face impossibility results anything like as bad as those for (Cond). Lewis (1976) showed how to make sense of conditionals defined by equations like (Cond*) by defining $\text{Prob}_\alpha$ in terms of an operation of *imaging* on α. However, (Cond*) too has contentious results. For example, since $\text{Prob}_\alpha(\neg\beta) = 1 - \text{Prob}_\alpha(\beta)$, $\text{Prob}(\alpha \mathrel{\Box\!\!\rightarrow} \neg\beta) = \text{Prob}_\alpha(\neg\beta) = 1 - \text{Prob}_\alpha(\beta) = 1 - \text{Prob}(\alpha \mathrel{\Box\!\!\rightarrow} \beta) = \text{Prob}(\neg(\alpha \mathrel{\Box\!\!\rightarrow} \beta)$.[21] But for possible α, by (Cond*):

$\text{Prob}((\alpha \mathrel{\Box\!\!\rightarrow} \beta) \wedge (\alpha \mathrel{\Box\!\!\rightarrow} \neg\beta)) = \text{Prob}(\alpha \mathrel{\Box\!\!\rightarrow} (\beta \wedge \neg\beta)) = \text{Prob}_\alpha(\beta \wedge \neg\beta) = 0$

By the standard probability axioms, the probability of a two-way disjunction is the sum of the probabilities of the disjuncts minus the probability of their conjunction, so we have a probabilistic version of the contentious principle of conditional excluded middle for possible antecedents:

(PCEM) $\text{Prob}((\alpha \mathrel{\Box\!\!\rightarrow} \beta) \vee (\alpha \mathrel{\Box\!\!\rightarrow} \neg\beta)) = 1$

Conditional excluded middle, the disjunction in (PCEM), is valid in Stalnaker's logic of conditionals, but invalid in Lewis's logic of counterfactuals.

Recently, natural language data have been used in defence of conditional excluded middle (Williams 2010). However, its epistemological consequences are highly problematic. Here is an example. Consider an epistemically reasonable probability distribution over worlds. Let *w* be a possible world where a fair coin is not tossed, *h* a possible world where it is tossed and comes up heads and *t* a possible world where it is tossed and comes up tails, *h* and *t* being as far as possible symmetrically related to *w*. Let *W* be true only in *w*, *H* true only in *h*, and *T* true only in *t*. Since *W* holds in only one world, for any α either *W* entails

[21] Compare Boethius' claim that to negate the consequent is to negate the conditional (Kneale and Kneale 1962: 191).

α in the probability space, in which case $W \mathrel{\Box\!\!\rightarrow} \alpha$ has probability 1, or $W$ is incompatible with α in that space, in which case $W \mathrel{\Box\!\!\rightarrow} \alpha$ has probability 0. In particular, therefore, $W \mathrel{\Box\!\!\rightarrow} ((H \vee T) \mathrel{\Box\!\!\rightarrow} H)$ has probability 1 or 0. Suppose that it has probability 1. Then, in effect, the distribution treats it as certain that *h* is selected over *t* as 'closer' to *w*: roughly, it is certain that in *w* if the coin had been tossed it would have come up heads. But that is epistemically quite unreasonable. The only alternative is for $W \mathrel{\Box\!\!\rightarrow} ((H \vee T) \mathrel{\Box\!\!\rightarrow} H)$ to have probability 0. But then $W \mathrel{\Box\!\!\rightarrow} \neg((H \vee T) \mathrel{\Box\!\!\rightarrow} H)$ has probability 1, so by conditional excluded middle $W \mathrel{\Box\!\!\rightarrow} ((H \vee T) \mathrel{\Box\!\!\rightarrow} \neg H)$ has probability 1, and so too has $W \mathrel{\Box\!\!\rightarrow} ((H \vee T) \mathrel{\Box\!\!\rightarrow} T)$. But that is equally epistemically unreasonable. Thus conditional excluded middle does not permit the required epistemic uncertainty between the worlds *h* and *t* with respect to *w*. Nor would it help to say that it is indeterminate (whatever that means) whether *h* or *t* is selected, and therefore indeterminate whether $W \mathrel{\Box\!\!\rightarrow} ((H \vee T) \mathrel{\Box\!\!\rightarrow} H)$ has epistemic probability 1 or 0, since that is not an available epistemic state for us. Notably, the problem does *not* generalize to Lewis's semantics, since he can straightforwardly treat *h* and *t* as equally close to *w* and so evaluate both $(H \vee T) \mathrel{\Box\!\!\rightarrow} H$ and $(H \vee T) \mathrel{\Box\!\!\rightarrow} T$ as false at *w*, and so assign probability 0 to both $W \mathrel{\Box\!\!\rightarrow} ((H \vee T) \mathrel{\Box\!\!\rightarrow} H)$ and $W \mathrel{\Box\!\!\rightarrow} ((H \vee T) \mathrel{\Box\!\!\rightarrow} T)$.[22]

What all these considerations suggest is that there is no reasonable way of fully vindicating the suppositional procedure. It breaks down for inconsistent antecedents, and may well do so for mundanely possible ones too. It does not follow that we do not use the suppositional procedure. It is plausibly our normal, unreflective way of evaluating counterfactual conditionals. But it is not 100% reliable. Like many of our methods of judgment, it is a useful but fallible heuristic. In particular, the linguistic data used to support the principle of conditional excluded middle may be the misleading outputs of such a heuristic.

For present purposes, the full power of the suppositional procedure is not needed. We need only consider a consequence of it that is independent of conditional excluded middle. Suppose that β is seen as inconsistent with γ. Then, normally, β is still seen as inconsistent with γ under the counterfactual supposition of α. Epistemically, the suppositional procedure treats combining β and γ under the counterfactual supposition of α as tantamount to combining $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ and $\alpha \mathrel{\Box\!\!\rightarrow} \gamma$. Thus an inconsistency in the former is seen as tantamount to an inconsistency in the latter. That yields the following heuristic as a by-product of the suppositional procedure:

(HCC) Given that β is inconsistent with γ, treat $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ as inconsistent with $\alpha \mathrel{\Box\!\!\rightarrow} \gamma$

Thus, recognizing that β is inconsistent with γ, if we accept $\alpha \mathrel{\Box\!\!\rightarrow} \beta$, we reject $\alpha \mathrel{\Box\!\!\rightarrow} \gamma$. We can apply this rule when drawing out the implications of any counterfactual supposition α. In practice, when using the suppositional procedure, we need not explicitly accept $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ in order to reject $\alpha \mathrel{\Box\!\!\rightarrow} \gamma$; it is enough to accept β within the counterfactual supposition of α, because the suppositional procedure treats those as equivalent.

For example, all this can be applied to the case of (13) and (14). 'Sick children in the mountains of South America at the time cared' is obviously incon-

[22] For those uncomfortable with doubly embedded subjunctive conditionals, one can replace the outer one by an ordinary strict conditional, and do the whole argument with $\Box(W \supset ((H \vee T) \mathrel{\Box\!\!\rightarrow} H)$ in place of $W \mathrel{\Box\!\!\rightarrow} ((H \vee T) \mathrel{\Box\!\!\rightarrow} H)$.

sistent with 'Sick children in the mountains of South America at the time did not care' (on the relevant readings), so in accordance with (HCC) we treat (13) as inconsistent with (14). Thus, having verified (14), we treat ourselves as having falsified (13). More economically, having rejected 'Sick children in the mountains of South America at the time cared' within the counterfactual supposition of the antecedent, we can directly reject (13) by the suppositional procedure. The upshot is the same. As already noted, the impossibility of the antecedent 'Hobbes (secretly) squared the circle' plays no role in our reasoning; logical relations under that supposition are treated no differently from usual.

For many purposes, we can consider a simpler heuristic in place of (HCC):

(HCC*) If you accept one of $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ and $\alpha \mathrel{\Box\!\!\rightarrow} \neg\beta$, reject the other.

(HCC*) has the advantage over (HCC) of not using 'inconsistent', a term which could do with some clarification. The suppositional procedure endorses (HCC*) as a heuristic: if you accept $\alpha \mathrel{\Box\!\!\rightarrow} \beta$, you accept $\beta$ under the counterfactual supposition of $\alpha$, so by normal reasoning you reject $\neg\beta$ under the counterfactual supposition of $\alpha$, so you reject $\alpha \mathrel{\Box\!\!\rightarrow} \beta$; likewise with $\beta$ and $\neg\beta$ interchanged.

(HCC) and (HCC*) are equivalent under normal conditions. First, start with (HCC). Clearly, $\beta$ is inconsistent with $\neg\beta$. Then (HCC) tells you to treat $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ as inconsistent with $\alpha \mathrel{\Box\!\!\rightarrow} \neg\beta$. Thus, if you accept one of them, you should reject the other. In other words, you should obey (HCC*). Conversely, start with (HCC*), and let it be given that $\beta$ is inconsistent with $\gamma$. Thus $\gamma$ entails $\neg\beta$. So, normally, from $\alpha \mathrel{\Box\!\!\rightarrow} \gamma$ you can derive $\alpha \mathrel{\Box\!\!\rightarrow} \neg\beta$, by an informal analogue of Deduction within Conditionals.[23] But (HCC*) tells you not to accept both $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ and $\alpha \mathrel{\Box\!\!\rightarrow} \neg\beta$. So, normally, you should not accept both $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ and $\alpha \mathrel{\Box\!\!\rightarrow} \gamma$. In other words, you should obey (HCC). However, since it is sometimes artificial to introduce an explicit negation when two sentences are obviously inconsistent, (HCC) may be the more natural heuristic.

There is psychological evidence that people reason in accordance with (HCC*), treating pairs of conditionals with the same antecedent and contradictory consequents as inconsistent, whether the conditionals are indicative or subjunctive (Evans, Handley and Over 2005: 1049-50). Moreover, if one evaluates the probabilities of counterfactuals according to (Cond*) (of which (Cond) is a special case), then $\mathrm{Prob}(\alpha \mathrel{\Box\!\!\rightarrow} \neg\beta) = 1 - \mathrm{Prob}(\alpha \mathrel{\Box\!\!\rightarrow} \beta)$, so if one of $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ and $\alpha \mathrel{\Box\!\!\rightarrow} \neg\beta$ is probable, the other is improbable, so not both can be accepted.

For the orthodox, (HCC) and (HCC*) are only heuristics because they lead you to reject true counterpossibles when $\alpha$ is impossible. However, it is plausible that usually, when counterfactual conditionals arise in practice, their antecedents are possible. In those cases, (HCC*) will never lead you astray. For if you accept one of $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ and $\alpha \mathrel{\Box\!\!\rightarrow} \neg\beta$, and it is true (which is not the responsibility of (HCC*)), (HCC*) will tell you to reject the other one, which will be false by (19). Given the near-equivalence of (HCC) and (HCC*), (HCC) will share much of the qualified reliability of (HCC*). Thus, on an orthodox logic of counterfactuals, both (HCC) and (HCC*) are reasonable though fallible heuristics.

[23] Even for the unorthodox, (DC) is *normally* reliable. Even for the orthodox, there are the abnormal cases discussed at the end of section 3, where (DC) must be restricted because the operative standard of validity is real world validity and the object-language contains fancy modal operators such as the rigidifying 'actually' operator. Such abnormal cases are rare in practice.

Even on an unorthodox view of counterfactuals, the prospects for fully vindicating (HCC) and (HCC*) are bleak. The previous problems for the suppositional procedure with inconsistent antecedents apply just as much to them. For example, consider counterpossibles with explicit contradictions as antecedents:

(61) $(\alpha \wedge \neg\alpha) \mathrel{\Box\!\!\rightarrow} \alpha$
(62) $(\alpha \wedge \neg\alpha) \mathrel{\Box\!\!\rightarrow} \neg\alpha$

Both (61) and (62) look highly plausible; surely conjunctions counterfactually imply their conjuncts. But if both (61) and (62) are true, then they are consistent, even though they have the same antecedent and inconsistent consequents. Now unorthodox theorists may reject some instances of (61) and (62), for instance when α itself is 'No conjunction counterfactually implies its conjuncts' or the like. But they cannot plausibly reject one of them in *all* cases. For instance, let α be 'The Liar is true', so $\alpha \wedge \neg\alpha$ makes the dialetheist claim about the Liar paradox that the Liar is both true and not true. Dialetheists both assert that the Liar is true *and* assert that the Liar is not true. Presumably, therefore, both (61) and (62) should hold on this reading of α, even for the unorthodox. Thus even they should regard (HCC) and (HCC*) as fallible heuristics, not as marking exceptionless rules of the logic of counterfactuals.

The situation is this. The naive evaluation of some counterpossibles, such as (13), involves a move from the rejection of the consequent under the counterfactual supposition of the antecedent to the rejection of the whole conditional. The principles that rationalize such a move turn out to hold only for the most part; they are fallible heuristics. There are independent reasons to expect them to fail for at least some counterpossibles. Thus we should not rely on them uncritically.

How should we expect the heuristic evaluation of $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ and $\alpha \mathrel{\Box\!\!\rightarrow} \neg\beta$ to go when α entails both β and ¬β? In effect, since both β and ¬β would eventually emerge as we developed the counterfactual supposition α for long enough, the heuristics make it a race between the contradictories as to which emerges first. If β emerges first, we accept $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ and reject $\alpha \mathrel{\Box\!\!\rightarrow} \neg\beta$ before ¬β has time to emerge. If ¬β emerges first, we accept $\alpha \mathrel{\Box\!\!\rightarrow} \neg\beta$ and reject $\alpha \mathrel{\Box\!\!\rightarrow} \beta$ before β has time to emerge. Proponents of impossible worlds misinterpret this computational, inferential difference in terms of the relative closeness of impossible $\alpha \wedge \beta$ and impossible $\alpha \wedge \neg\beta$ worlds.

One advantage of the heuristics account is that it explains our inattention to the impossibility of the antecedent in our cognitive processing of many counterpossibles.[24] By contrast, accounts such as Brogaard and Salerno's that postulate a special standard of relative closeness for impossible worlds, apparently quite different from that appropriate for possible worlds, fail to explain the lack of felt adjustment to such a special standard in our cognitive processing of counterpossibles.

[24] In some cases, we do attend to the impossibility of the antecedent. For example, we prefer 'If Hobbes had squared the circle, he would have done something geometrically impossible' to 'If Hobbes had squared the circle, he would have done nothing geometrically impossible'. The heuristic account has no difficulty with such examples: cognitively, the connection from antecedent to consequent is still easier with the former counterfactual than with the latter.

We can test the three heuristics—the suppositional procedure, (HCC), and (HCC*)—by trying them out on three examples offered by Cian Dorr as "some manifestly false counterfactuals whose antecedents seem to be metaphysically impossible" (Dorr 2008: 37):[25]

(63) If I were a dolphin, I would have arms and legs.

(64) If it were necessary that there be donkeys, it would be impossible that there be cows.

(65) If there were unicorns, none of them would have horns.

Let us grant Dorr that in each case the antecedent is indeed metaphysically impossible.

First, consider (63). Let α = 'I am a dolphin', β = 'I have arms and legs', and γ = 'I do not have arms and legs'. In developing the antecedent supposition 'I am a dolphin', one holds fixed one's general knowledge of dolphins, including one's knowledge that they do not have arms and legs, and so derives 'I do not have arms and legs'. One thus accepts γ. Since β is manifestly inconsistent with γ, one rejects β. Any of the three heuristics then leads one to accept $\alpha \mathrel{\Box\!\!\rightarrow} \gamma$ and reject $\alpha \mathrel{\Box\!\!\rightarrow} \beta$, which is (63). Thus the employment of any of the heuristics explains our rejection of (63).

Next, consider (64). Let α = 'It is necessary that there are donkeys', β = 'It is impossible that there are cows', and γ = 'It is possible that there are cows' (where the modal operators are read non-epistemically, as Dorr intends). In developing the antecedent supposition 'It is necessary that there are donkeys' with an eye to (64), one considers the modal status of 'There are cows'. In actuality, there are cows, so cows are possible. The supposed necessity of donkeys presents no good reason not to hold those facts fixed. Indeed, if one treats the modal status of cows and donkeys on a par, one would have to make cows necessary too, and therefore still possible. Hence one still favours 'It is possible that there are cows'. One thus accepts γ. Since β is manifestly inconsistent with γ, one rejects β. Any of the three heuristics then leads one to accept $\alpha \mathrel{\Box\!\!\rightarrow} \gamma$ and reject $\alpha \mathrel{\Box\!\!\rightarrow} \beta$, which is (64). Thus the employment of any of the heuristics explains our rejection of (64).

Finally, consider (65). Let α = 'There are unicorns', β = 'No unicorn has horns', and γ = 'There are unicorns with horns'. In developing the antecedent supposition 'There are unicorns', one holds fixed with one's background conception of unicorns, including 'Unicorns have horns', and so derives 'There are unicorns with horns'. One thus accepts γ. Since β is manifestly inconsistent with γ, one rejects β. Any of the three heuristics then leads one to accept $\alpha \mathrel{\Box\!\!\rightarrow} \gamma$ and reject $\alpha \mathrel{\Box\!\!\rightarrow} \beta$, which is (65). Thus the employment of any of the heuristics explains our rejection of (65).

Our impression that Dorr's examples are all false can be explained by our reliance on a fallible heuristic. On the suggested explanations, in no case does the impossibility of the antecedent figure in our assessment. We simply do not consider the modal status of the antecedent. That seems to fit the phenomenology of unreflectively assessing (63)-(65).

Any of the heuristics can also be applied to mathematical examples:

(66) If 289 were divisible by 3, 290 would also be divisible by 3.

[25] For (65), Dorr cites Kripke's influential argument for the impossibility of unicorns (Kripke 1980: 156).

A natural impression is that (66) is false. The explanation is along the usual lines. Let α = '289 is divisible by 3', β = '290 is divisible by 3', and γ = '290 is not divisible by 3'. In developing the antecedent supposition '289 is divisible by 3', one holds fixed one's background knowledge that the successor of a multiple of 3 is never a multiple of 3, and so derives '290 is not divisible by 3'. One thus accepts γ. Since β is manifestly inconsistent with γ, one rejects β. Any of the three heuristics then leads one to accept $\alpha \mathrel{\Box\!\!\rightarrow} \gamma$ and reject $\alpha \mathrel{\Box\!\!\rightarrow} \beta$, which is (66). Thus the employment of any of the heuristics explains our rejection of (66). But of course there are also equally cogent mathematical deductions of '290 is divisible by 3' from '289 is divisible by 3'; they are just a bit longer and less psychologically salient. Rejecting (66) would be unfaithful to the natural use of counterpossibles in formulating mathematical proofs.

Any of the three heuristics leads us to the false conclusion that (66) is false, and in very similar ways to the conclusions that (63)-(65) are false. To say the least, we should be wary of those conclusions too. Our impression that (63)-(65) are 'manifestly false' may well the product of our reliance on a highly fallible heuristic.[26]

Of course, we are not completely helpless victims of our heuristics. Through conscious theoretical reflection, we can sometimes inhibit their operation. Our mastery of reasoning by *reductio ad absurdum* in mathematics shows our ability to defeat (HCC), (HCC*), and the suppositional procedure. For example, we accept both the counterpossibles (58) and (59) in the proof that there is no largest prime, even though they have the same antecedent and mutually inconsistent consequents. Even in less formal settings, it is not psychologically compulsory to call off the search for β amongst the counterfactual consequences of α once ¬β has turned up. If we are asked an open-ended question such as 'What would have been the consequences if α had been the case?', we can continue the search in a way that allows for mutually inconsistent counterfactual consequences to emerge. That is in effect what we do when asked 'Could α have obtained?' (Williamson 2007: 162). Nevertheless, despite our ability to inhibit their operation, heuristics remain the default, to which we may always be liable to revert when off our guard. For instance, if one puts aside one's mathematical sophistication, it is not hard to feel that (58) and (59) are mutually inconsistent after all.[27]

A useful analogy, already suggested in section 2, is with our naïve reactions to true universal quantifications with empty subject terms:

(67) Every dolphin in Oxford has arms and legs.

[26] Our practical use of 'If I were you' counterpossibles and the associated imaginative exercises may also be explicable in terms of such heuristics, without appeal to any special semantics of the *de se* or the like.

[27] An alternative hypothesis is that our behaviour manifests sensitivity to conversational implicatures rather than use of heuristics. For example, in asserting (13) one might be thought to counterfactually imply that sick children in the mountains of South America kept track of Hobbes's activities. If such a line worked, it would provide an alternative defence of orthodoxy about counterpossibles. However, attempts to cancel the alleged conversational implicatures work poorly: for example, saying 'I don't mean to imply that those children were aware of Hobbes' hardly dissolves the cognitive dissonance caused by asserting 'If Hobbes had squared the circle, sick children in the mountains of South America at the time would have cared'. Although conversational implicatures may sometimes be in the mix, they are not the crucial ingredient.

(68) Every unicorn is hornless.

A natural inclination is to judge (67) and (68) false. Even when one is told that there are no dolphins in Oxford and no unicorns, one still feels some resistance to accepting (67) and (68) respectively. That resistance is explicable by the hypothesis that we accept (69) and (70) on the basis of background information about dolphins and unicorns respectively, and are then inclined to reject (67) as inconsistent with (69) and (68) as inconsistent with (70):

(69) Every dolphin in Oxford lacks arms and legs.

(70) Every unicorn has a horn.

That suggests heuristics for universal quantification analogous to (HCC) and (HCC*):

(HUQ) Given that φ is inconsistent with ψ, treat 'Every σ φs' as inconsistent with 'Every σ ψs'.

(HUQ*) If you accept one of 'Every σ φs' and 'Every σ ¬φs', reject the other.

On the standard semantics for the universal quantifier, (HUQ) and (HUQ*) go extensionally wrong when and only when σ is empty in extension.[28]

We can come to recognize the limitations of (HUQ) and (HUQ*) through natural reasoning. For instance, suppose that our rejection of (67) leads us to accept its negation:

(71) Not every dolphin in Oxford has arms and legs.

From (71) we can validly reason to (72):

(72) Some dolphin in Oxford lacks arms and legs.

From (72) we can in turn validly reason to (73):

(73) There is a dolphin in Oxford.

But we know (73) to be false. That may lead us to realize that (67) is not false, though its utterance may induce a false presupposition. (HUQ) and (HUQ*) are fallible heuristics, defeasible by theoretical reflection, but they are still our default.

A more general underlying cognitive pattern may explain these heuristics. For example, it is plausible that we use analogues of them for indicative as well as subjunctive conditionals, and for generic as well as universal quantifiers. We ignore the issue of the empty case. We continue using heuristics that do so even when the empty case is obviously relevant, until we resort to conscious reflection. Indeed, we may tend to use suppositional reasoning in evaluating universal and generic generalizations as well as conditionals. For instance, when asked to evaluate (67) or its generic analogue ('Dolphins in Oxford have arms and legs'), we may suppose that something is a dolphin in Oxford, and ask ourselves whether it has arms and legs.

Our theoretical grasp of universal quantification is currently more secure than it is of counterfactuals conditionals. We are consequently more comfortable in overruling (HUQ) and (HUQ*) than in overruling (HCC) and (HCC*). But it was not always so. Centuries of confusion about the existential import or otherwise of the universal quantifier bear witness to the difficulty of achieving an accurate view of the truth-conditions of sentences of our native language

[28] An example where they go wrong: 'Every σ which is a non-σ is a σ' and 'Every σ which is a non-σ is a non-σ'.

formed using the most basic logical constants.[29] Those who take themselves to have provided clear examples of false counterpossibles may be in a similar position to traditional logicians who took themselves to have provided clear examples of false universal generalizations with empty subject terms. Indeed, the primitively compelling nature of heuristics such as (HUQ) and (HUQ*) may have been the main obstacle to achieving a clear view of the truth-conditions of universal generalizations.

Imagine a philosopher attempting to craft a semantics for the universal quantifier to vindicate the heuristically driven judgments that (67) and (68) are false while (69) and (70) are true. He may invest immense patience and ingenuity in his project, but it is not going to end well. We should be similarly wary of attempts to craft a semantics for the counterfactual conditional to vindicate the heuristically driven judgments that some counterpossibles are false while others are true. There is a danger in semantics of unintentionally laundering cognitive biases into veridical insights, a danger evident in the semantics of generics, where some theories make sentences such as 'Muslims are terrorists' come out true.[30]

In the case of the universal quantifier, proper understanding was finally achieved through systematic, highly general semantic and logical theorizing, rather than by a more data-driven approach. The same may well hold for the counterfactual conditional. At any rate, it is methodologically naïve to take the debate over counterpossibles to be settled by some supposed examples of clearly false counterpossibles. As we have seen, a simple and mostly reliable heuristic would lead us to judge them false even if they were true.

On the view developed here, our assessments of counterfactuals are often based on fallible heuristics such as (HCC), (HCC*), or the suppositional procedure. How far should that view make us sceptical more generally about reliance on pre-theoretic assessments of counterfactuals in philosophy, semantics and elsewhere? Several points are worth noting.

First, the heuristics are reliable over wide ranges of cases. Just as we can gain lots of perceptual knowledge by relying on perceptual heuristics that are reliable over wide ranges of cases but fail under special conditions, so we can gain lots of modal knowledge by relying on heuristics for counterfactuals. Blanket scepticism is not a sensible response.

Second, the problems posed for the heuristics by counterpossibles concern the rejection of $\Box\rightarrow$ conditionals, not their acceptance. Arguably, the key judgments in thought experiments, for instance that in such-and-such a Gettier case the subject would not know, involve the acceptance of $\Box\rightarrow$ conditionals (Williamson 2007: 179-207).

Third, nothing said here impugns the reliability of counterfactual judgments made on the basis of mathematical reasoning.

Fourth, the heuristics at issue are of such a general and pre-reflective nature that one might conjecture them to be universal amongst humans, more or less hard-wired into us independently of intelligence and education, with relatively little individual variation. But of course that does not mean that there will be no such variation in the inputs (acceptance of one counterfactual) or corresponding-

[29] See Peters and Westerståhl (2006: 124-27) for a concise defence of the modern view of existential import. For a contrasting view see Strawson (1952: 163-79).
[30] Sterken (2015) makes this point.

ly in the outputs (rejection of another). Furthermore, whether the individual takes the output of the heuristic at face value or second-guesses it on the basis of theoretical reflection may well be highly sensitive to individual variation and educational background. It is an appropriate locus for the application of philosophical expertise.

Fifth, even without theoretical reflection, we can inhibit the operation of the heuristics, as we sometimes need to do in order to maintain consistency. For instance, as already noted, we can continue the imaginative search for counterfactual consequences of a subjunctive supposition in an open-minded way that allows contradictions to arise. The operation of one heuristic may also be preempted or inhibited by the operation of another.

Finally, the theorist who overrides the heuristic in favour of more reflective considerations should expect to feel some residual unease, at least at first. No matter how cogent the reflective considerations, the heuristic is too stupid to understand them; instead, it just goes on blindly pressing to have its way. If our access to the logic and semantics of our own language is essentially mediated by fallible heuristics, true theories may always feel Procrustean to us.[31]

## References

Berto, F., French, R., Priest, G. and Ripley, D. 2016, "Williamson on Counterpossibles", Typescript.

Brogaard, B. and Salerno, J. 2013, "Remarks on Counterpossibles", *Synthese*, 190, 639-60.

Carnap, R. 1947, *Meaning and Necessity: A Study in Semantics and Modal Logic*, Chicago: Chicago University Press.

Davies, M. and Humberstone, L. 1980, "Two Notions of Necessity', *Philosophical Studies*, 38, 1-30.

Dorr, C. 2008, "There Are No Abstract Objects", in Sider, T., Hawthorne, J. and Zimmerman, D. (eds.), *Contemporary Debates in Metaphysics*, Oxford: Blackwell, 32-63.

Edgington, D. 2008, "Counterfactuals", *Proceedings of the Aristotelian Society*, 108, 1-21.

Evans, J., Handley, S. and Over, D. 2005, "Suppositions, Extensionality, and Conditionals: A Critique of the Mental Model Theory of Johnson-Laird and Byrne (2002)", *Psychological Review*, 112, 1040-52.

Evans, J. and Over, D. 2004, *If*, Oxford: Oxford University Press.

[31] This paper is based on my 2015 Beth Lecture in Amsterdam. Other versions of this material were presented at the University of Oxford, University of Edinburgh, the University of Connecticut, Yale University, Ohio State University, a conference on conditionals at the University of Belgrade, a conference on the impossible at the University of Turin, and a Mind, World and Action course in Dubrovnik. I thank the audiences and also Jennifer Nagel, Carola Barbero, Andrea Iacona, and Alberto Voltolini for many helpful questions, objections, and comments. Thanks also to Francesco Berto, Rohan French, Graham Priest, and David Ripley for showing me their critique of an earlier version of this paper (Berto, French, Priest, and Ripley 2016), which has prompted some important amplifications of the discussion of heuristics. An earlier version of some of the present material appeared as Williamson 201X.

Field, H. 1980, *Science Without Numbers: A Defence of Nominalism*, Oxford: Blackwell.

Field, H. 1989, *Realism, Mathematics, and Modality*, Oxford: Blackwell.

Fine, K. 1994, "Essence and Modality", *Philosophical Perspectives*, 8, 1-16.

von Fintel, K. 1998, "The Presupposition of Subjunctive Conditionals", in Sauerland, U. and Percus, O. (eds.), *The Interpretive Tract*, MIT Working Papers in Linguistics 25, Cambridge (MA): MIT Press, 29-44.

Grice, P. 1989, *Studies in the Ways of Words*, Cambridge (Ma): Harvard University Press.

Hájek, A. 2011, "Triviality Pursuit", *Topoi*, 30, 3-15.

Kment, B. 2014, *Modality and Explanatory Reasoning*, Oxford: Oxford University Press.

Kratzer, A. 2012, *Modals and Conditionals*, Oxford: Oxford University Press.

Kneale, W. and Kneale, M. 1962, *The Development of Logic*. Oxford: Clarendon Press.

Kripke, S. 1980, *Naming and Necessity*, Oxford: Blackwell.

Lewis, D. 1970, "General Semantics", *Synthese*, 22, 18-67.

Lewis, D. 1973, *Counterfactuals*, Oxford: Blackwell. Page references to 2nd ed., 1986.

Lewis, D. 1976, "Probabilities of Conditionals and Conditional Probabilities", *Philosophical Review*, 85, 297-315.

Nolan, D. 1997, "Impossible Worlds: A Modest Approach", *Notre Dame Journal for Formal Logic*, 38, 535-72.

Peters, S. and Westerståhl, D. 2006, *Quantifiers in Language and Logic*, Oxford: Clarendon Press.

Salmon, N. 1989; "The Logic of What Might Have Been", *Philosophical Review*, 98, 3-34.

Stalnaker, R. 1968, "A Theory of Conditionals", *American Philosophical Quarterly Monographs*, 2, 98-112.

Sterken, R. 2015, "Generics, Content and Cognitive Bias", *Analytic Philosophy*, 56, 75-93.

Strawson, P. 1952, *Introduction to Logical Theory*, London: Methuen.

Vetter, B. 2016, "Williamsonian Modal Epistemology, Possibility Based", *Canadian Journal of Philosophy*, 46, 766-95; repr. in McCullagh, M. and Yli-Vakkuri, J. (eds.), *Williamson on Modality*, London: Routledge 2017.

Williams, J. R. G. 2010, "Defending Conditional Excluded Middle", *Noûs*, 44, 650-68.

Williamson, T. 2006, "Indicative Versus Subjunctive Conditionals, Congruential Versus Non-Hyperintensional Contexts", in Sosa, E. and Villanueva, E. (eds.), *Philosophical Issues*, *Volume 16: Philosophy of Language*, Oxford: Blackwell, 310-33.

Williamson, T. 2007, *The Philosophy of Philosophy*, Oxford: Blackwell.

Williamson, T. 2016, "Modal Science", *Canadian Journal of Philosophy*, 46, 453-92; repr. in McCullagh, M. and Yli-Vakkuri, J. (eds.), *Williamson on Modality*, London: Routledge 2017.

Williamson, T. 201X, "Counterpossibles", *Topoi*, forthcoming.

# Impossible Worlds and the Intensional Sense of 'and'

*Luis Estrada-González*

*Institute for Philosophical Research, UNAM*

*Abstract*

In this paper I show that the 'and' in an argument like Lewis' against concrete impossible worlds cannot be simply assumed to be extensional. An allegedly 'and'-free argument against impossible worlds employing an alternative definition of 'contradiction' can be presented, but besides falling prey of the usual objections to the negation involved in it, such 'and'-free argument is not quite so since it still needs some sort of premise-binding, thus intensional 'and' is needed and that suffices to block the argument at a stage prior to the steps about negation.

*Keywords:* Intensional 'and', contradiction, impossible worlds, 'not', Lewis

## 1. Introduction

Lewis' argument against impossible worlds (Lewis 1986: 7) goes as follows. If worlds are concrete entities and the expression 'at world *w*' works as a restricting modifier—that is, if it restricts the quantifiers within its scope to parts of *w*—, then it should distribute through the extensional connectives.[1] Let us say that a modifier M distributes over an n-ary connective © if and only if, if $M(©(\phi 1,\ldots, \phi n)$, then $©(M(\phi 1),\ldots, M(\phi n))$. This means in particular that 'At *w*, both *A* and not *A*', where such a world *w* is

[1] Consider a sequence of sentences $\phi_1,\ldots, \phi_n$, another sentence $\psi$, and a valuation *v* that takes values in a collection *V* of values. An *n*-ary connective © is *extensional* if and only if the following condition is satisfied for any *v*:

$$\text{If } v(\phi_i) = v(\psi) \text{ then } v(©(\phi_i)) = v(©(\phi_1,\ldots, \phi_{i-1}, \psi, \phi_{i+1},\ldots, \phi_n)).$$

© is *truth-functional* if and only if $v(©(\phi_i))$ is a function from $V^n$ to *V*. Lewis speaks of truth-functional connectives, which is a rather strong requisite: truth-functionality implies extensionality (see Humberstone 2011: Ch. 3). The opposition truth-functional/intensional, very frequent in the literature, is not problematic in some cases. However, I will employ the proper opposition extensional/intensional.

*Argumenta* 2, 2 (2017): 227-239
ISSN 2465-2334


DOI 10.23811/43.arg2017.gon

called then an *impossible world*, entails 'At *w*, A, and it is not the case that at *w*, A'. Thus, any contradiction at some world turns into an overt inconsistency at the actual world. But *pace* the dialetheist, there are no contradictions at the actual world, so there are no such impossible worlds.

This argument has been widely studied and virtually all its parts challenged, mostly trying to save impossible worlds for their theoretical applicability or even indispensability (see Nolan 1997: Sect. 2), or because they should stand along with possible worlds just by parity of reasoning (e.g. Vander Laan 1997, Beall and van Fraassen 2003). One option has been to reject Lewis' concretism. Thus, some think that worlds, whether possible or impossible, are abstract entities (e.g. Mares 1997, Vander Laan 1997). Others accept concretism for possible worlds but not for impossible worlds (e.g. Berto 2010, Divers 2002: Ch. 5). Others think that worlds different from the actual are not concrete but are not abstract either, at least in the usual sense of the term (e.g. Zalta 1997). Also, conceptions of logical (or, rather, modal) space radically different from Lewis' have been put forward (see Yagisawa 2010). All these are, however, realist views on worlds, so finally there are fully anti-realist views of impossible worlds (see Nolan 1997, Beall 2008).

As to other attempts to block Lewis' argument, some authors have maintained that Lewis begs the question against impossibilists in describing the nature of the modifier 'at *w*' (e.g. Lycan 1994). There is also the charge that using the law of non-contradiction in this context is question-begging (e.g. Yagisawa 1988) and, finally, those who would try to swallow the overt inconsistency at the actual world (e.g. Yagisawa 1988; Priest 2006a for several other reasons).

I argue here that the extensionality of 'and' in Lewis' original argument and almost all the subsequent discussion on it is an assumption that can be coherently challenged and rejected. Note that this is not a "desperate" move to block Lewis' argument: As mentioned above, there are already plenty of options in the market. I just want to emphasize that there is another option available, not necessarily in conflict with the others; that such option questions an early step of Lewis' argument, usually taken as safe; and that such option is worth pondering because of its implications for other logical, linguistic and metaphysical topics.

The plan of the paper is as follows. In Section 2 I show that reasons independent of the question about impossible worlds prevent from simply assuming that 'and' is extensional in Lewis' argument because 'and' is in general intensional. The next sections are devoted to answer some possible objections, in increasing degree of seriousness, against this appeal to the intensional sense of 'and' to block Lewis' argument. In Section 3 the charges of change of subject and that Lewis' argument holds at least for a sense of 'and' are dealt with. In Section 4 a supposedly 'and'-free argument against concrete impossible worlds is presented, which is still subject to already well-known objections to the extensionality of 'not', and I show that the reasons to support the intensional 'not' that blocks that argument belong to the family of reasons to support the intensional 'and' of Section 2. In Section 5 I argue that there is no reason to stick only to intensional 'not' although it suffices to block both arguments against concrete impossible worlds and, more importantly, that such 'and'-free argument is not quite so: Intensional 'and' is needed as a premise-binder

and once properly analyzed the argument is blocked before the dubious steps concerning 'not'.

## 2. The Intensional Sense of 'and'

Lewis in his original argument against impossible worlds, and virtually all the authors discussing it, uncritically assumed the extensionality of 'and' so that the modifier 'at *w*' passes through it; or, more exactly, that 'and' satisfies the following property:

*Distribution of M over 'and':* For every $A$, $B$ and restricting modifier $M$,

$$\frac{M(A \text{ and } B)}{M(A) \text{ and } M(B)}$$

Some inferences involving 'and', notoriously 'and'-elimination, have been recognized as problematic in this context, but almost always derivatively—see, e.g. Priest (2008: 172), Kiourti (2010: Ch. 4)—and, for reasons to be discussed in Sections 4 and 5, the extensionality of 'and' is ultimately left untouched.[2] Lewis (1982: 102ff) also considered failures of 'and'-introduction and 'and'-elimination for indexes that are nothing ontologically heavy but "(fragments of) corpuses of information". This comes as a "more conservative" approach to sentences that might be both true and false after his famous "dogmatic" rejection of sentences that are both true and false simpliciter. But in this context, such dogmatic rejection amounts to say that the indexes in which sentences are both true and false cannot be concrete worlds, but that is precisely what is at stake.

However, a sense of 'and' that does not satisfy *Distribution of M over 'and'* has been introduced and defended on grounds other than contradictions and impossible worlds: Fusion, also called sometimes 'intensional', 'group-theoretical', or 'multiplicative' conjunction in the tradition of substructural logics, which serves to express the idea that the joint content of the conjuncts is needed to entail something. Humberstone (2011, esp. Ch. 5) has criticized the introduction of new connectives, especially for different senses of 'and', on the basis that most of the times matters pragmatic are confused with matters semantic and no real logical need for a new 'and' is put forward. But the picture for fusion seems promising. If an indicative conditional is required to connect contents (semantic, informational or whatnot) beyond those expressible by truth values alone, as would happen if it embodies in an object language a suitable notion of entailment, then a conditional like 'If I like sandwiches then this is an amazing journal.' might not be true even if both antecedent and consequent are true. But Read (1981; 2003) has argued that if an indicative conditional connecting more than truth values is true, then the 'and' of an inference like

$$\frac{\text{Both } A \text{ and } B}{\text{So if } A \text{ then } B}$$

[2] With the notable exceptions of Nolan 1997 and Yagisawa 2010, to be discussed at the end of this section.

cannot be an extensional connective, for the premise could be true and the conclusion false.

Read (1981) has also claimed that phrasings like 'It cannot be both *A* and not *B*' do not express an impossibility but rather the inferential connection between '*A*' and '*B*'. In that case, 'It cannot be both *A* and not *B*' would imply 'If *A* then *B*', so 'It cannot be both that the premises of an argument are true and the conclusion false' would imply 'If the premises of an argument are true then the conclusion is true'. In general, this 'and' in the very characterization of validity should be non-extensional, as would be the 'and' and other premise-binding devices in an argument.[3]

Such interaction between indicative conditionals and 'and' has appeared independently in the debate about impossible worlds. Let us suppose that an impossible world is more generally a world where logic is different from the actual world, not only one where sentences of the form 'Both *A* and not *A*' are true. In particular, sentences of the form 'If *A* then *A*' may fail to be true. Such a conditional still comes with a residual, $\circ$ (the double bar indicates that the inference goes both ways):

$$\frac{A \text{ implies that if } B \text{ then } C}{A \circ B \text{ implies } C}$$

i.e. a sort of conjunction, just read 'and' for $\circ$ , which nonetheless should be in general not extensional to keep company the failure of 'If *A* then *A*'.[4]

Fusion is a very general proposition-binder, and especially a premise-binder, that can work as the formal counterpart of that 'and'.[5] Every proposition *A* is evaluated at an index *i*, which is a (possibly empty) *multiset* of conditions of evaluation additional to its truth value. A *binding*, *fusion* or *intensional conjunction* of the propositions *A* and *B*, denoted here *'A∘B'*, at an index *i* has the same content as the multi-union in *i* of the content of *A* at some index *j* and the content of *B* at some index *k,* both of the latter related to *i*.[6] More perspicuously,

[3] The intensional 'and' has also served to express relevance as a necessary component of validity understood as truth-preservation rather than a condition to be added to it; see Read 2003. Moreover, appealing to it can be helpful to deal with paradoxes like Curry or the so-called 'paradoxes of validity', as in Priest 2015.

[4] For more details, see Slaney 1990. That this conditional comes with such a residual goes against Priest's (2008: 172) idea that if an impossible world is a world where the laws of logic are different (from those of a certain designated world, presumably the actual one, where presumably classical logic holds), then the behavior of conditionals, and only of them, change at impossible worlds, because conditionals express the laws of logic: If conditionals are affected, conjunctions are affected as well, and so it is not true that "[t]here are no [non-normal, impossible] worlds at which $A \wedge B$ is true, but *A* is not".

[5] In what follows I stick closely to Mares' simple and intuitive explanation of fusion (cf. Mares 2012).

[6] A multiset is like a set, except that it may contain identical elements repeated a finite number of times. If *X* and *Y* are multisets then, for every element *a*, if it appears *n* times in *X* and *m* times in *Y*, then it appears *n+m* times in the multi-union of *X* and *Y*.

$$\frac{A \text{ and } B}{A,\ B}$$

is not valid, because sentences in general not only have a truth value, but additional content provided by the indexes in which they are uttered or evaluated. The more general form of the inference above helps to show why it is not valid:

$$\frac{(A_j \text{ and } B_k)_i}{A_{ji},\ B_{ki}}$$

Thus, from $(A_j \text{ and } B_k)_i$ one cannot infer $A_i$ (or $B_i$), but at most $A_{ji}$ (or $B_{ki}$).[7]

In classical zero-order logic, for all practical purposes there is but only one index, the empty index, so $i = j = k$ and the contents bound in that index are present separately in that index and that validates the usual rule of conjunction-elimination. Start with the general version

$$\frac{(A_j \text{ and } B_k)_i}{A_{ji},\ B_{ki}}$$

Since $i = j = k$,

$$\frac{(A_i \text{ and } B_i)_i}{Aii,\ Bii}$$

hence

$$\frac{(A \text{ and } B)}{A,\ B}$$

because the multi-union of empty multisets is empty, $ii = i$, all indexes are identical and hence can be obviated.

Nonetheless, the more general version

$$\frac{\text{At } i,\ A \text{ and } B}{\text{At } i\ A \text{ and at } i,\ B}$$

is not valid, because what there is at $i$ is the multi-union of contents that may be present only at indexes different from $i$. When the indexes are worlds, even concrete worlds like Lewis', *Distribution of M over 'and'* needs not hold.

[7] Here is a rather simple example. Suppose that $A$ holds at $j$, which is the multiset of conditions of evaluation [b, c, c, d], and that $B$ holds at $k$, which is the multiset of conditions of evaluation [a, a, c]. Then '$A$ and $B$' holds at the index $i$ = [a, a, b, c, c, c, d], but it is not true that $A$ (or $B$) alone holds at [a, a, b, c, c, c, d]. Indeed, one of the conditions in [a, a, c] (respectively, [b, c, c, d]) might prevent having $A$ (respectively, $B$) alone, as in the case of connexive logic mentioned below.

Thus, given that indicative conditionals in general are not extensional—or, at the very least, in general they are not extensional in impossible worlds—, 'and' in general is not extensional, either. But if 'and' in general is not extensional, *Distribution of M over 'and'* needs not hold. Also, if 'and' in general is not extensional, there is no reason to suppose that it is in a sentence of the form 'A and not A' which is true (by hypothesis) at an impossible world. And if there is no reason to suppose that 'and' is extensional in an impossible world, then 'at w' does not necessarily distribute through it. But if it does not, Lewis' original argument against (concrete) impossible worlds does not run.

Someone might retort that, even if all this formal account connecting conjunction and the conditional is sound, no actual example of the failure of 'and'-elimination has been provided. Of course, that 'and'-elimination is invalid does not mean that every instance of it should be rejected. No doubt 'There is a laptop on the desk' validly follows from "There are a laptop and a tablet on the desk', but if even a single instance of 'and'-elimination can be found in which the conclusion fails to follow from the premises, then that will be sufficient to show that 'and'-elimination is not valid.

In fact, there are at least three classes of such instances. The first is provided by the cancellation theory of negation (see Priest 2006: 31ff for an overview). According to it, an assertion says one thing, and its denial withdraws it. Thus, a contradiction cancels its content out and leaves nothing behind. Therefore, nothing follows from a contradiction, in particular, none of its conjuncts. The second class is provided by inferences of the following form, familiar from connexive logic (cf. Thompson 1991):

$$\frac{A\text{, and } A \text{ does not follow from } A}{A}$$

But this means that A follows from *A* even under the condition that it does not. Thus, the inference from '*A* and *B*' to '*A*' is rejected on the grounds that *'B'* might assert something that would countermand the inference from *'A'* to *'A'*. Finally, consider this counterexample introduced by Gillian Russell (in preparation). Let us say that a literal, i.e. an atomic proposition or its negation, is solo if and only if it is not embedded in a conjunction. Then the following inference seems to go from true premises to a false conclusion

$$\frac{\text{This very sentence is not SOLO and snow is white}}{\text{This very sentence is not SOLO.}}$$

In all fairness, at least Nolan (1997) and Yagisawa (2010) have regarded *Distribution of M over 'and'*, as well as other classical inferences connecting 'and' and conditionals discussed above, problematic when it comes to impossible worlds. The distinctiveness of my approach is twofold. First, I relate the discussion explicitly to Lewis' argument against impossible worlds, whereas those principles are just sam-

ples in Nolan's broader discussion of what should go in an 'ultralogic' to reason about any kind of situation, including impossible worlds. Second, Nolan thinks that the inferences might fail to hold only at very strange situations, whereas I regard the counterexamples to those inferences rather natural, at least as natural as are those for the intensional conditional. Yagisawa (2010: 183f) makes the failure of *Distribution of M over 'and'* a feature only of impossible worlds, almost as if 'and' could have arbitrary truth-conditions at an impossible world, but although the idea that impossible worlds are semantically arbitrary at least for some connectives is very widespread—see for example the reports in Beall (2009)—, such arbitrariness is not necessary to make sense of the failure of certain classical inferences concerning 'and', what is needed is just a bit of generalization.[8]

## 3. Changing the Subject and Two Kinds of 'and' for Impossible Worlds

It can be objected that the approach above is merely an *ignoratio elenchi*, that Lewis' was discussing extensional 'and' and that his argument indeed works for it. I think this is wrong even as an exegetical point. Lewis argues that if sentences of the form 'A and not A' are true at some worlds, then they are true at the actual world, which would be wrong and then the hypothesis should be rejected. In doing that he simply assumed that connectives are extensional, that is, Lewis does not argue that there are no concrete impossible worlds for some specific formalizations of sentences of the form '*A* and not *A*', but that there are no concrete impossible worlds, period. I think this is the usual underlying dialectical approach to Lewis' argument. For example, when Lycan (1994) or Kiourti (2010) challenge the idea that 'at w' passes through 'not', they do not think of themselves nor are taken by their interlocutors as changing the subject of Lewis argument, but as providing a better understanding of the logic of the sentences of the form '*A* and not *A*' and its consequences for the status of impossible worlds. The same goes for the analysis of 'and'.

The second objection strengthens the previous one and considers the case of a language suitable to talk about impossible worlds with both kinds of 'and': Even if

[8] The appeal to an intensional 'and' has consequences for other discussions in the vicinity. In a broader discussion of the appropriate semantic clauses for possibility and impossibility once impossible worlds are admitted, Divers (2002: Ch. 5) says that the usual clause for impossibilities

(IP) *A* is impossible if and only if there is no world, *w*, such that at *w*, *A*

would be false, for impossibilities hold at impossible worlds, and that amending (IP) in the following way

(IP*) *A* is impossible if and only if there is an impossible world, *w*, such that at *w*, *A*

would be useless: Suppose $A \wedge \neg A$ holds at an impossible world *w*. Then presumably each conjunct holds at *w*, that is, at *w*, *A*, and at *w*, $\neg A$. But since each conjunct holds at an impossible world, each of them is impossible, by (IP*), although clearly in general they are not. Again, it is presupposed that the conjunction is extensional; if it is not, this argument against (IP*) does not work. I am not saying that (IP*) is the right semantic clause for impossibility; what the proper stance about the semantic clauses for possibility and impossibility should be is a more general problem that I will not discuss here.

Lewis' argument does not work for contradictions with an intensional 'and', it still works for contradictions in which 'and' is extensional, so having both kinds of 'and' at an impossible world would turn at least some contradictions there into contradictions at the actual world. Two interrelated answers can be given to this objection. The first one is that in the case described, Lewis' argument would only show that there are no concrete impossible worlds with both kinds of 'and', not that there are no concrete impossible worlds at all. An argument against all concrete impossible worlds would require an argument for the necessary presence of both kinds of 'and' in the language corresponding to an impossible world. In the absence of such argument for necessarily having both kinds of 'and', that Lewisian argument would be precisely the argument for the inadmissibility of the extensional 'and' in a language for impossible worlds.

The second reply is that an argument for the inadmissibility of the extensional 'and' in languages for impossible worlds could be an excessively strong demand. What would be needed is at most an argument for the inappropriateness of the extensional 'and' to represent sentences of the form '*A* and not *A*'. This is an idea not uncommon among thinkers of contradictions. In several places of *Metaphysics* Aristotle attributed to the Heracliteans the thesis that 'All contradictions (and only them) are true' and reconstructed their views as follows.[9] Everything is in state of flux at every moment, so and a thing cannot be described truly to be an *F* because it would be to fix it, and the same considerations are made for not being an F, but it can be described truly and fully as being an *F* as well as not being an *F*. Hence all contradictions, but none of their components separately, are true, because it would be the only way to capture the changing nature of things (cf. *Metaphysics* $1005^{b}25$, $1007^{b}26$, $1012^{a}25$). More recently, when discussing how contradictions could be observed, Paul Kabay (2010: 110) proposes that contradictions are different from other conjunctions, that they are more like "single whole entities", not a "unified structure with distinct parts", and then he claims that contradictions, if true, would involve the same thing in all the same respects, as Aristotle stated, so there would be nothing to separate. Actually, Priest (2006: 11) also notes that Aristotle in Prior Analytics 57b3 cannot accept 'and'-elimination for contradictions since he claims that "contradictories cannot both entail the same thing". Finally, that the conjunction of contradictions involved in Hegel's dialectic and theses like that of the unity and identity of opposites is different from ordinary conjunctions was raised by some commentators (e.g. Wetter 1958; Havas 1981; Priest 1989: 397).

A final objection would say again that, in spite of what the thinkers above could argue, there is a change of subject, because sentences of the form 'A and not A', where 'and' is intensional, are not real contradictions; real contradictions are sentences of the form 'A and not A' where 'and' is extensional, and Lewis' is arguing against those. I have pointed out in the preceding section that this is exegetically

[9] Priest sometimes conflates trivialism—'Everything is true'—with a version of Heracliteanism—'All contradictions are true'—(cf. Priest 2007: 131), although sometimes he acknowledges that the identification depends on certain assumptions, notoriously 'and'-elimination (see Priest 2006: 56). Aristotle too thought that semantic Heracliteanism could be equated with trivialism, but he was more cautious; cf. *Metaphysics* $1012^{a}25$.

incorrect. But, moreover, I have showed above that extensional 'and' can be seen as a limit, extremely idealized case of the general case but where there is but one, dispensable index. Further argumentation is needed to claim that only conjunctions (and negations) uttered or evaluated at a certain single index are the "real" conjunctions and negations, and so the only connectives that can produce "real" contradictions.

Thus, provided the correctness of the independent arguments for the idea that 'and' is in general intensional and the views of thinkers of contradictions just mentioned, what needs to be argued for is the idea that sentences of the form '*A* and not *A*' are correctly formalized using an extensional 'and'—not to mention the stronger claim that only it provides the correct formalization of those sentences—even if it is available in the language.

## 4. Another Definition of 'Contradiction' and the Intensionality of 'not'

Even if Lewis uses the definition of a contradiction as a sentence of the form 'A and not A' and it is the way it has usually been discussed, an apparently 'and'-free argument can be formulated taking the definition as a pair of sentences '*A*, not *A*'.[10] In that case, the appeal to the intensional 'and' would seem useless, for the rule

For every *A*, *B* and restricting modifier *M*,

*Distribution of M over 'and'*:
$$\frac{M(A, B)}{M(A), M(B)}$$

applied to contradictions as pairs

$$\frac{M(A, \text{not } A)}{M(A), M(\text{not } A)}$$

sounds good: If we have a pair of contradictories at M, it sounds plausible that we have at M each of the members of the pair, unlike the case of contradictions of the form '*A* and not *A*'.

However, this move leaves Lewis' argument at the mercy of already known objections to others of his assumptions. If the hypothesis of this kind of impossible worlds is taken seriously, there is no reason to grant that 'at w' passes through 'not'

[10] Stalnaker's version of the argument against impossible worlds uses the latter option (cf. Stalnaker 1996/2002). Lewis (1986) does not consider contradictions of the form '*A*, not *A*'. This is not accidental. His discussion of *ex falso quodlibet* in Lewis (1982: 104) makes clear that he considered that the pair version may be susceptible to counterexamples because the separate premises *A*, not *A* might track contents from different indexes, while he was confident that the single premise '*A* and not *A*' could avoid that. However, as discussed in Section 2, the intensional 'and' also tracks contents from different indexes.

to produce the contradiction at the actual world, i.e. that 'not' satisfies the following property (see Kiourti 2010: Ch. 4):

For every $A$ and restricting modifier $M$,

*Distribution*[11] *of M over 'not'*:

$$\frac{M(\text{not } A)}{\text{Not } (M\ A)}$$

This feature of 'not' is as related to conditionals as are the motivations for the intensional 'and'. If an indicative conditional is required to connect contents beyond those expressible by truth values alone, some indicative conditionals of the forms 'If $A$ and not $A$ then $B$' and 'If $A$ then $B$ or not $B$' are not true. That might be achieved by an intensional negation of A, denoted here '$\ominus A$'.[12] $\ominus A$ is true at an index of evaluation i if and only if A is false at some index j maximal with respect to all the indexes compatible with i. More perspicuously, in the more general version

$$\frac{\text{At } i, \text{ not } A}{\text{Not at } i, A}$$

is not valid, because the truth of 'Not $A$' at i is the untruth of $A$ at an index $k$ that may be different from i. (I am sure the reader can give the explicit version of this as was done for 'and'.) When the indexes are worlds, even concrete worlds like Lewis', *Distribution of M over 'not'* needs not hold. Remember that in classical logic, for all practical purposes there is but only one index, so $\ominus A$ is true if and only if A is untrue at that single index.

## 5. Intensional 'and' Strikes Back

The possibility of using the intensional 'not' to block Lewis' argument raises a further objection to the approach advocating the intensional sense of 'and'. Since 'not' appears in both arguments, is not it "the only relevant" logical notion, as Stalnaker's (1996/2002: 58) Louis would say? And, moreover, if intensional 'not' suffices to block the arguments against concrete impossible worlds with either definition of contradiction, should not the changes be kept at minimum and avoid introducing

[11] Some people might feel that this is not really a case of distribution over a connective, but rather a sort of commutativity or other phenomenon. However, unary connectives like negation can be present in special cases of distribution as it was defined right in Section 1.

[12] I will follow Restall's explanation of this negation (cf. Restall 1999). The required notions are as follows: Suppose that there are accessibility relations between indexes, one of which constitutes a partial order and that can be denoted '$\leq$'. That an index $i$ is *maximal* with respect to an index $j$ means that for every index $x$, if $j \leq x$ then $x \leq i$. What the compatibility of $j$ with $i$ exactly means depends on substantial philosophical ideas, but at least the following characterization might be admitted without much problems: $j$ is *compatible* with $i$ if (1) $Rij$, (2) there is another relation between $i$ and $j$, denoted '$Cij$', such that $C$ is symmetric, non-reflexive and that, for all indexes $x$ and $y$, if $Cij$, $x \leq i$ and $y \leq j$, then $Cxy$.

the intensional 'and' which would serve at most for one version? I can think of at least two answers to these questions. First, if the arguments for intensional 'and' from the nature of indicative conditionals reconstructed in Section 2 are right and if defining a contradiction as a sentence of the form '*A* and not *A*' is on equal footing with defining it as a certain kind of pair, intensional 'and' can be legitimately invoked when in presence of contradictions of the mentioned form, as in Lewis' original argument and most of the subsequent discussion.[13]

But the most important reason for not sticking only to the intensional 'not' is that the claims that intensional 'and' is irrelevant for *Distribution of M over ','*, or that 'not' is the only relevant logical notion in the 'and'-free argument, are contentious, for the argument depends on the way of interpreting bound premises. Again, if indicative conditionals connect contents beyond those expressible by truth values alone, and if in one's logic logical consequence and the conditional are related by a "deduction theorem", that is, *B* is a logical consequence from $A_1,\ldots, A_n$ if and only if the corresponding conditional 'If $A_1$ and… and $A_n$ then *B*' is a theorem, those 'and's must be in general intensional to keep company the indicative conditional; see Read (1988: Ch. 3); Slaney (1990). But that would mean that also the commas in '$A_1,\ldots, A_n$' are intensional premise-binders.[14] Thus, even if the argument against concrete impossible worlds started 'At *w*, *A*, not *A*', there are reasons to think that a modifier like 'at w' does not pass through comma or any other premise-binder. The intensional premise-binding has thus an occurrence in the argument prior to that of the intensional 'not' and has to be analyzed first. But, once the premise-binder is properly analyzed, Lewis argument can be resisted at an earlier stage.

## 6. Conclusion

Intensional 'and' represents another option in the market to block Lewis' argument against impossible worlds, whether under the assumption that contradictions are sentences of the form '*A* and not *A*' or under the alternative definition of contradictions as certain pairs, as in Stalnaker's version of the argument. It not only does not compete with the well-known strategy of being more precise about 'not' in the context of impossible worlds, but together with this it grist to the mill of those who

[13] Kiourti (2010: Ch. 4), building upon some concerns expressed by Nolan (1997), considers two possible problems that might constrain the number of admissible intensional notions. The first one is that the intensional truth-conditions of connectives "no longer [would] allow us to break these down to their individual components at such worlds, treating them instead as atomic predicates of worlds." This "might seem extreme", as she says, but her own answer, which seems right, is that it seems less extreme if one notes that this is a feature of impossible worlds that not necessarily spills over into any other worlds. The second worry, which she considers more pressing and leads her to adopt just the intensional 'not', is that one might fall short of principles to reason about an entire ontology comprising both possible and impossible worlds. But that is not the case; intensional connectives are disassociated from certain inferences but not from all of them.

[14] Like fusion, this feature of the comma has been widely studied in the so-called substructural logics; see for example Restall 2000, Ch. 2.

think that it is possible to have an extended modal realism *á la* Yagisawa without contradictions at the actual world, by blocking Lewis' argument at an earlier step. The remarks about 'and' would also serve to block one of the most ancient arguments against true contradictions, the argument for explosion, popularized by another Lewis. The most common strategies are either rejecting Disjunctive Syllogism or appealing to an intensional sense of 'or' (see Read 1988: Ch. 2). But if there is no reason to suppose that an 'and', especially that of a contradiction, is extensional, or if premise-binding is tighter than usually thought, the argument can be blocked already at the step following the assumption(s).

The appearance of intensional 'and' in the context of impossible worlds invites one to think about further implications of this notion for the philosophy of logic, especially when it comes to the understanding of some connectives and on a possible debate about the preferability of one definition of 'contradiction' over the other.[15]

## References

Aristotle, *Metaphysics*, in Barnes, J. (ed.), *The Complete Works, Volume 2: The Revised Oxford Translation*, Princeton: Princeton University Press, 1984.

Beall, J.C. 2008, "Transparent Disquotationalism", in *Deflationism and Paradox*, Beall, J.C. and Armour-Garb, B. (eds.), Oxford: Oxford University Press, 7-22.

Beall, J.C. and van Fraassen, B. 2003, *Possibilities and Paradox. An Introduction to Modal and Many-Valued Logic*, Oxford: Oxford University Press.

Berto, F. 2010, "Impossible Worlds and Propositions: Against the Parity Thesis", *The Philosophical Quarterly*, 60, 471-86.

Divers, J. 2002, *Possible Worlds*, London-New York: Routledge.

Havas, K. 1981, "Some Remarks on an Attempt at Formalising Dialectical Logic", *Studies in Soviet Thought*, 22, 257-64.

Kabay, P. 2010, *On the Plenitude of Truth. A Defense of Trivialism*, Lambert Academic Publishing.

Kiourti, I. 2010, *Real Impossible Worlds: The Bounds of Possibility*. PhD thesis, University of St Andrews.

[15] This paper was written under the support from the European Union through the European Social Fund (Mobilitas grant no. MJD310), the PAPIIT project IA401015 "Tras las consecuencias. Una visión universalista de la lógica (I)", and the CONACyT project CCB 2011 166502 "Aspectos filosóficos de la modalidad". I would like to thank Franz Berto, Edwin Mares, Chris Mortensen, Daniel Nolan, Graham Priest, Stephen Read and Carlos Romero for their helpful comments on earlier versions of the paper. I am also grateful to the members of the Seminario de Filosofía del Lenguaje, UNAM, especially to Axel Barceló, Delia Belleri, Maite Ezcurdia, Mario Gómez-Torrente and Alessandro Torza for their thorough reading of a previous version and helpful discussion. Finally, I would like to thank the two anonymous reviewers for their careful reading and comments.

Lewis, D. 1982, "Logic for Equivocators", in *Papers in Philosophical Logic*, Cambridge: Cambridge University Press, 97-110. Originally appeared in *Noûs*, 16, 431-41.

Lewis, D. 1986, *On the Plurality of Worlds*, Oxford: Blackwell.

Lycan, W. 1994, *Modality and Meaning*, Dordrecht: Kluwer.

Mares, E. 1997, "Who's Afraid of Impossible Worlds?", *Notre Dame Journal of Formal Logic*, 38, 516-26.

Mares, E. 2012, "Relevance and Conjunction", *Journal of Logic and Computation*, 22, 7-21.

Nolan, D. 1997, "Impossible Worlds: A Modest Approach", *Notre Dame Journal of Formal Logic*, 38, 535-72.

Priest, G. 1989, "Dialectic and Dialetheic", *Science & Society*, 53, 388-415.

Priest, G. 2006, *Doubt Truth to Be Liar*, Oxford: Oxford Clarendon Press.

Priest, G. 2006a, *In Contradiction*, second edition, Oxford: Oxford University Press.

Priest, G. 2007, "Paraconsistency and Dialetheism", in Gabbay, D. and Woods, J. (eds.), *The Many Valued and Nonmonotonic Turn in Logic*, Handbook of the History of Logic, vol. 8, Amsterdam: Elsevier, 129-204.

Priest, G. 2008, *An Introduction to Non-Classical Logic. From If to Is*, second edition, Cambridge: Cambridge University Press.

Priest, G. 2015, "Fusion and Confusion", *Topoi* , 34, 55-61.

Read, S. 1981, "Validity and the Intensional Sense of 'and'", *Australasian Journal of Philosophy*, 59, 301-307.

Read, S. 1988, *Relevant Logic: A Philosophical Examination of Inference*, Oxford: Basil Blackwell.

Read, S. 2003, "Logical Consequence as Truth-Preservation", *Logique et Analyse*, 183-184, 479-93.

Rescher, N. and Brandom, R.B. 1980, *The Logic of Inconsistency: A Study in Non-Standard Possible Worlds Semantics and Ontology*, Oxford: Basil Blackwell.

Restall, G. 1999, "Negation in Relevant Logics: How I Stopped Worrying and Learned to Love the Routley Star", in Gabbay, D. and Wansing, H. (ed.), *What is Negation?*, Dordrecht: Kluwer Academic Publishers, 53-76.

Russell, G. in preparation, "Could There Be *No* Logic?".

Slaney, J. 1990, "A General Logic", *Australasian Journal of Philosophy*, 68, 74-88.

Stalnaker, R. 1996/2002, "Impossibilities", in *Ways a World Might Be. Metaphysical and Anti-Metaphysical Essays*, 2003, Oxford: Oxford University Press, 55-67.

Thompson, B.E.R. 1991, "Why is Conjunctive Simplification Invalid?", *Notre Dame Journal of Formal Logic*, 32 (2), 248-54.

Vander Laan, D. 1997, "The Ontology of Impossible Worlds", *Notre Dame Journal of Formal Logic*, 38, 597-620.

Wetter, G.A. 1958, *Dialectical Materialism*, London: Routledge and Kegan Paul.

Yagisawa, T. 1988, "Beyond Possible Worlds", *Philosophical Studies*, 53, 175-204.

Zalta, E. 1997, "A Classically-Based Theory of Impossible Worlds", *Notre Dame Journal of Formal Logic*, 38, 640-60.

# S4 to 5D

*Takashi Yagisawa*

*California State University, Northridge*

*Abstract*

The modal logical axiom 4 is widely accepted. It is the characteristic axiom of the modal logical system S4, which is subsumed under the most popular modal logical system S5. Axiom 4 is equivalent to $\Diamond\Diamond P \rightarrow \Diamond P$ ("If possibly possibly $P$, then possibly $P$"), which requires that the accessibility relation between worlds be transitive.

There is a powerful argument (Hugh Chandler 1976, Nathan Salmon 1981, 1989) against axiom 4. It rests on the thought that an ordinary object could have had a slightly different origin from its actual origin but could not have had an origin very different from its actual origin. By constructing a sorites-like sequence of possible worlds at which the origin of a given object shifts incrementally along the sequence, the argument concludes that accessibility is not transitive, i.e. that what is possibly possible may not be possible.

A recent attempt to defend S4 from this argument (Murray and Wilson 2012) proposes that we abandon the absolute notion of possibility and instead accept a world-indexed notion of possibility; each world comes with its own version of possibility.

I offer a different defense of S4, which preserves both axiom 4 and the absoluteness of possibility. Its key move is to postulate objects as extended not only in physical space-time but in logical space as well, that is, as "five-dimensional" worms. Since S4 and the absolute notion of possibility are very intuitive, quite useful, and widely well regarded, and since my proposal saves both of them, I take the proposal to constitute an argument in favor of "five-dimensionalism."

*Keywords*: Modality, S4, Origin Essentialism, Five-Dimensionalism, Impossible World.

## 1. Introduction

S5 is the most popular modal logical system among modal metaphysicians. S4 is weaker than S5. So, anyone who accepts S5 should also accept S4. But there is trouble with S4. Or so argue Hugh S. Chandler and Nathan Salmon. Chandler's argument is directed against Alvin Plantinga's claim that nothing is possible at some possible worlds and not possible at others. Chandler aims to establish "that what is possible

*Argumenta* 2,2 (2017): 241-261
ISSN 2465-2334


DOI 10.23811/44.arg2017.yag

varies from world to world."[1] Salmon takes Chandler's argument and elaborates on it more broadly as an argument against S4.[2] Even though it is Salmon, not Chandler, who explicitly targets S4, I shall be concerned with Chandler's original version of the argument for its simplicity. The general thrust of my discussion applies equally well to Salmon's version.

Chandler's argument threatens the characteristic axiom of S4, namely:

Axiom 4: $\Diamond\Diamond P \rightarrow \Diamond P$.

It says that whatever is possibly possible is possible.[3] Independently of commitment to stronger S5, this axiom seems well worth saving by itself. Here are two examples illustrating its plausibility:

(i) I have no child but could have had one. If I had a child, that child could have had a child. So, I could have had a grandchild.

(ii) There is a physical particle which does not split but could have split into two particles. If it had so split, each of the two resulting particles could also have split into two particles, producing four further particles in total. So, there could have been four particles instead of just one.

These are just examples and do not amount to an argument in favor of Axiom 4, but their overwhelming natural plausibility should strongly encourage us to attempt search for a way to save Axiom 4 from any objection against it. That is the spirit in which I approach Chandler's argument.

That spirit is shared by Adam Murray and Jessica Wilson, who propose a way to save Axiom 4.[4] Their rescue attempt, however, comes at a serious cost and also seems ineffective. I wish to propose a different way to save Axiom 4 without the cost and with effectiveness.

## 2. Preliminaries

According to standard modal logic, truth is indexed to a world, and truth of a possibility at a world is truth at an accessible world:

(PT): $\Diamond P$ is true at a world $w$ if and only if P is true at some world accessible from $w$.

In general accessibility is any dyadic relation between worlds, but given (PT), Axiom 4 constrains it to be transitive: for any worlds $w$ and $w'$, if $w'$ is accessible from some world that is accessible from $w$, then $w'$ is accessible from $w$. Intuitively, accessibility is intended to be relative possibility: $w'$ is accessible from $w$ if and only if all that holds at $w'$ is possible relative to $w$. This is strictly just an intuitive idea, for if it were taken seriously as a definition of accessibility, possibility would be defined in terms of relative possibility, and the latter would remain in need of further definition if we wanted

[1] Chandler 1976: 106.
[2] Salmon 1989.
[3] An equivalent formulation of Axiom 4 is: $\Box P \rightarrow \Box\Box P$ (whatever is necessary is necessarily necessary).
[4] Murray and Wilson 2012.

to complete definitions of all modal notions. (I am assuming that relative possibility is a modal notion.) This will become important when we evaluate the proposal by Murray and Wilson.

The possibility operator $\Diamond$ may be interpreted in different ways: logical possibility, metaphysical possibility, physical possibility, human psychological possibility, legal possibility (for a given society), etc. Chandler focuses on metaphysical possibility, the subject matter of Plantinga's claim.[5] Salmon's elaboration on Chandler's argument and the criticism by Murray and Wilson do not deviate from this focus. My discussion will also be strictly about metaphysical possibility, and no other kind of possibility will be considered in this paper.

## 3. Argument Against Axiom 4

A bicycle is an artifact which is manufactured out of, or originates from, many parts. If a particular bicycle in fact originated from particular parts, then that very same bicycle could possibly have originated from the same particular parts except for one spoke; in place of that spoke, a completely different spoke might have been used to manufacture a bicycle numerically identical with the original.

Some might wish to deny this and insist that even if all other original parts were used in the same way as for the original bicycle, if one spoke were different, then the resulting bicycle would not be numerically identical with the original bicycle. If such obstinate essentialism concerning origin is accepted, Chandler's argument is blocked at the outset.[6] It is not my intention to block Chandler's argument this way; neither is it the intention of Murray and Wilson. If a spoke would make a difference to the numerical identity of the resulting bicycle, there seems to be no principled reason to deny that a small part of a spoke would also make a difference. But if so, it seems that a large molecule would make a difference, too. But if a molecule would, why not an atom? And it seems implausible to insist that one atomic difference in origin would destroy the numerical identity of the manufactured bicycle.

On the other hand, if no original parts had been used to manufacture a bicycle except for one original spoke, then the resulting bicycle would not have been numerically identical with the original bicycle. It is unclear how many of the original parts should have been used to retain the numerical identity of the original bicycle. To avoid deciding this tangential issue with a bicycle, or any other familiar kind of object, and simplify discussion, Chandler conjures up an imaginary kind of object, which he calls *alpha*. He stipulates that any object of this kind—any alpha—originates from three parts and that it is possible for any alpha that in fact originated from three particular parts to have originated from two of those parts plus a different third part, but not from one of those parts plus two different parts, or from none of those parts. Thus, alphas are compound material objects with a very unusual condition for origin: for any alpha $x$, if $x$ originated from matter $m$, then $x$ could not have originated from matter two-thirds or more different from $m$. No familiar compound material object

[5] Or possibility "in a broad logical sense"; see Chandler 1976: 106.
[6] The phrase "obstinate essentialism" is due to Salmon.

has such a simple and sharp condition for origin. This unfamiliar nature of alphas should not discourage us from going along with Chandler's scenario. Far from it, we should welcome the simplification Chandler brings forth by the introduction of alphas, as a measure that helps us focus our attention squarely on the core issue of the transitivity of accessibility without distraction.[7]

With this simplifying assumption, Chandler considers a particular alpha, which he calls *Alfred*. Let us say that at a world $w_0$ Alfred exists and originated from matter consisting of three particular parts, 1-2-3. So at $w_0$ Alfred could have originated from 4-2-3, where 4 is distinct from 1.[8] That is, at some world $w_1$ accessible from $w_0$, Alfred exists and originated from 4-2-3. Since at $w_1$ Alfred originated from these three parts, at $w_1$ Alfred could have originated from two of them plus a new part, say, from 4-5-3, where 5 is distinct from 2 and from 1. That is, at some world $w_2$ accessible from $w_1$, Alfred exists and originated from 4-5-3.

Suppose for *reductio* that $w_2$ is accessible from $w_0$. Then Alfred exists and originated from 4-5-3 at a world accessible from $w_0$, which means that at $w_0$ Alfred could have originated from 4-5-3. But being an alpha and having originated from 1-2-3 at $w_0$, Alfred at $w_0$ could not have originated from 4-5-3. A contradiction! Therefore, $w_2$ is inaccessible from $w_0$. Since $w_2$ is accessible from $w_1$, which is accessible from $w_0$, accessibility is not transitive. This is Chandler's argument.

Let us put Chandler's argument in a regimented way to reveal its logical structure:

P: Alfred originated from 4-5-3.
1. P is not true at any world accessible from $w_0$.
2. P is not possibly true at $w_0$.
3. P is true at $w_2$, and $w_2$ is accessible from $w_1$, which is accessible from $w_0$.
4. P is possibly possibly true at $w_0$.
5. Some proposition, viz., P, is not possibly true but possibly possibly true, at $w_0$.
6. Accessibility is not transitive, i.e., Axiom 4 is false.

1 and 3 are true, 2 follows from 1, 4 follows from 3, 5 follows from 2 and 4, and 6 follows from 5. Or so claims Chandler.

It is important to note that the alpha which exists at $w_2$ and originated from 4-5-3 is supposed to be indeed Alfred, and not some other alpha. If it were some alpha other than Alfred, then that alpha's having originated from 4-5-3 would not make $w_2$ inaccessible from $w_0$. It is perfectly possible at $w_0$ that some alpha other than Alfred exists and originated from 4-5-3.

[7] To see clearly that vagueness in the condition of origin is largely a distraction, observe that Chandler's argument can be easily adapted according to a scenario incorporating vagueness as long as the vagueness permits the starting point and the end point in the *sorites*-like series of minutely shifting origin of a given object, to be an uncontroversially possible case of origination and an uncontroversially impossible case of origination, respectively.

[8] It goes without saying that 4 is also distinct from 2 and 3. Similar suppositions will not be noted explicitly henceforth.

It is also important to bear clearly in mind that only one kind of possibility, viz., metaphysical possibility, is in question and that therefore only one accessibility relation figures in the scenario used in the argument. It is intended that the accessibility relation holding between $w_1$ and $w_0$ is the very same accessibility relation holding between $w_2$ and $w_1$, and this very same accessibility relation is argued to fail between $w_2$ and $w_0$. If this were not so, the argument would not succeed in exhibiting a single non-transitive accessibility relation.

## 4. Attempt to Save Axiom 4

Murray and Wilson do not consider Chandler's Alfred example but discuss Salmon's different example instead. Their response to Salmon's version of Chandler-ish argument, however, can be adapted straightforwardly to become a response to Chandler's original argument.

Murray and Wilson in effect object to Premise 3 in Chandler's argument above. Their position is that 3 is ambiguous, for "accessible" is used ambiguously. And this ambiguity is inherited by 4 and 5 so that 6 does not follow. They deny that one single accessibility relation figures in the scenario and deny that one single notion of possibility is iterated in 4 or 5; that is how they propose to block the argument.

They phrase this move in terms of their own technical notion, *considered as indicatively actual*. When $w_0$ is considered as indicatively actual, Alfred could have originated from 4-2-3, but not from 4-5-3, whereas when $w_1$ is considered as indicatively actual, Alfred could have originated from 4-5-3. Murray and Wilson consider themselves as borrowing this technical notion from two-dimensional semantics and using it not so much for semantic or epistemic purposes as for metaphysical purposes. They propose that metaphysical possibility is relative to a world considered as indicatively actual. If I understand their proposal correctly, this means, as I have already indicated, that they wish to block the argument by denying that one single accessibility relation is involved throughout the scenario. We may put it this way: When we adopt the point of view of $w_0$ at the beginning and say that Alfred could have originated from such-and-such parts, we are resorting to one accessibility relation, which we might call *accessibility*$_0$, and say that $w_1$ is accessible$_0$ from $w_0$. When we move on to adopt the point of view of $w_1$, however, we shift to a different accessibility relation, *accessibility*$_1$, and say that $w_2$ is accessible$_1$ from $w_1$. And when we say that Alfred at $w_2$ is impossible, we are returning to the point of view of $w_0$ and saying that $w_2$ is inaccessible$_0$ from $w_0$. There is no absolute accessibility common to the point of view of an arbitrary world, but there are only world-relative accessibility relations, one for each world (considered as indicatively actual). To each such accessibility relation corresponds a distinct notion of possibility: possibility$_0$ to accessibility$_0$, and possibility$_1$ to accessibility$_1$.

Here is how Murray and Wilson would see Chandler's argument:

P: Alfred originated from 4-5-3.
1'. P is not true at any world accessible$_0$ from $w_0$.
2'. P is not possibly$_0$ true at $w_0$.
3'. P is true at $w_2$, and $w_2$ is accessible$_1$ from $w_1$, which is accessible$_0$ from $w_0$.

4'. P is possibly$_0$ possibly$_1$ true at $w_0$.
5'. Some proposition, viz., P, is not possibly$_0$ true but possibly$_0$ possibly$_1$ true, at $w_0$.
6'. Accessibility is not transitive, i.e., Axiom 4 is false.

Clearly, 6' does not follow from 5'. Moreover, according to Murray and Wilson, it is not that P is possibly$_0$ possibly$_0$ true at $w_0$, it is not that P is possibly$_1$ possibly$_1$ true at $w_0$, and it is not that P is possibly$_k$ possibly$_k$ true at $w_0$, for any k; so there is no counterexample to Axiom 4 in Chandler's scenario.

Murray and Wilson's move to preserve Axiom 4 in the face of Chandler's argument is a radical move. It proliferates accessibility relations for metaphysical possibility, hence it proliferates varieties of possibility all of which fall under the umbrella notion of metaphysical possibility. Indeed, it produces as many varieties of metaphysical possibility as there are worlds to be considered as indicatively actual. But this seems undesirable. Metaphysical possibility is a kind of possibility among many different kinds of possibility, but it seems that there is only one kind of possibility that is metaphysical possibility, and it certainly seems that there are not as many varieties of metaphysical possibility as there are worlds (eligible to be considered as indicatively actual).

As we saw in the opening paragraph, Chandler claims against Plantinga that what is metaphysically possible "varies from world to world." Murray and Wilson are presumably on Plantinga's side but they claim that metaphysical possibility itself varies from world to world; once a particular variety of metaphysical possibility is fixed, what is possible does not vary from world to world, but what variety of metaphysical possibility is in question to begin with does so vary. This hardly seems like much of a defense of Plantinga, or Axiom 4.

Also, even if there are many different varieties of metaphysical possibility and corresponding accessibility relations, Chandler's argument seems to go through unscathed after all. Suppose with Murray and Wilson that metaphysical possibility is always relative to a world (considered as indicatively actual). Take the metaphysical possibility relative to the actual world $w_0$ (by considering $w_0$ as indicatively actual). This determines a particular accessibility relation, $R$. Everyone agrees that $w_1$ is $R$ to $w_0$. How about $w_2$ and $w_1$? Is $w_2$ $R$ to $w_1$? Alfred originated from 4-2-3 at $w_1$ and originated from 4-5-3 at $w_2$. Since Alfred is an alpha at $w_0$, $R$ is such that a world at which Alfred originated from matter that is two-thirds identical with the matter from which Alfred originated at a given world is $R$ to that given world. So $w_2$ is $R$ to $w_1$.

Do not be fooled into thinking that $w_2$ is not $R$ to $w_1$ on the ground that at $w_2$ Alfred originated from the matter only one-third identical with the matter from which Alfred originated at $w_0$. Considering $w_0$ as indicatively actual only determines the variety of metaphysical possibility, i.e., only fixes the accessibility relation to be $R$. It does not make the original matter of Alfred at $w_0$ be the object of comparison to the original matter of Alfred at $w_2$ when considering the question whether $w_2$ is $R$ to $w_1$.

Since the question is whether $R$ holds between $w_2$ and $w_1$, not whether $R$ holds between $w_2$ and $w_0$, the object of comparison should be the original matter of Alfred at $w_1$.[9]

At the same time, since Alfred's original matter at $w_2$ is only one-third identical with Alfred's original matter at $w_0$, $w_2$ is not $R$ to $w_0$. Thus, $w_1$ is $R$ to $w_0$, $w_2$ is $R$ to $w_1$, but $w_2$ is not $R$ to $w_0$. Therefore, $R$ is not transitive.

Although I am not in favor of Murray and Wilson's application of two-dimensional semantics according to which each world gives rise to a different variety of metaphysical possibility, there seems to be something attractive about their radical approach. We can use what I take to be the underlying spirit in which they offer the core idea of *considering a world as indicatively actual*. As we move from $w_0$ to $w_1$, or from $w_1$ to $w_2$, when considering how Alfred could or could not have originated, something metaphysical shifts. What shifts is not the variety of possibility, but reference. Or so I claim. In my view, a proper response to Chandler's argument will not only preserve Axiom 4 but also give us an opportunity to learn about how we refer to objects in modal space.

## 5. Five-Dimensionalism

Suppose that an elephant is standing calmly in an enclosure at a zoo. Call the elephant "Elfie." When a blind person touches Elfie's trunk, she is touching Elfie. She is touching Elfie by touching its trunk. When another blind person touches Elfie's belly, he is touching Elfie. He is touching Elfie by touching its belly. The two blind people touch one and the same elephant, Elfie, in two different places. When the first person says, "The animal I am touching is like a snake," she is speaking of Elfie and saying of Elfie that it is like a snake. When the second person says, "The animal I am touching is like a wall," he is speaking of Elfie and saying of Elfie that it is like a wall.

Two points should be noted for our purposes with this version of a familiar Indian parable. First, the two blind people are touching two different parts of Elfie. They are perceiving (by touch) two different objects, a trunk and a belly. There is certainly a sense in which they are perceiving one and the same object, namely, Elfie. But it is equally certainly the case that they are perceiving different parts of Elfie, and it is this fact that we should note well for our purposes. Second, the two blind people do not say, "What I am touching is like …" Instead they use the concept *animal* and say, "The animal I am touching is like …" This gives unity to the subject matter of their discourse. Both speakers are speaking of one and the same animal, viz., Elfie. Elfie has two different parts, and the two blind people are touching them, but these parts are not animals; they are animal parts. Thus, the two people are perceiving two different objects (as well as perceiving one animal) and speaking of one and the same common object (the animal), where the two objects are different parts of the common object.

[9] More cautiously put, if the original matter of Alfred at $w_0$ has some relevance to answering the question whether $w_2$ is $R$ to $w_1$, it is not clear what it is.

I propose to extend this picture to our consideration of modal space. At $w_0$, a person $p_0$ perceives a particular alpha, Alfred, points to it, and says, "The alpha I am pointing to originated from 1-2-3, but could have originated from 4-2-3." At $w_1$, a person $p_1$ perceives the same particular alpha, Alfred, points to it, and says, "The alpha I am pointing to originated from 4-2-3, but could have originated from 4-5-3." In the elephant parable, the two blind people are separated from each other in physical space. They are at two different locations and are perceiving two different spatial parts Elfie has at these locations. Here, $p_0$ and $p_1$ are separated from each other in modal space. They are at two different worlds and perceiving two different worldly parts (world stages) which Alfred has at these worlds. This presupposes, of course, that Alfred is extended in modal space, having different worldly parts at different worlds, in a way analogous to the way Elfie is extended in physical space, having different spatial parts at different locations. This picture may be said to be a picture of *five-dimensionalism*. It incorporates three physical spatial dimensions, one temporal dimension, and one modal dimension.[10] Embrace this five-dimensionalist way of understanding Chandler's scenario, and Axiom 4 is saved. Or so I claim.

It should be noted that even though $p_0$ at $w_0$ and $p_1$ at $w_1$ perceive the same particular alpha, namely Alfred, this does not preclude $p_1$'s also perceiving a different alpha (or $p_0$'s also perceiving a different alpha, for that matter). The worldly part Alfred has at $w_1$ may be a worldly part ($w_1$-stage) of another alpha; two different alphas may share one and the same $w_1$-stage. If this is indeed the case, then when $p_1$ says, "The alpha I am pointing to originated …, but could have originated …," $p_1$ is pointing to a particular $w_1$-stage shared by two different alphas. So $p_1$'s use of the definite description "the alpha I am pointing to" has no unique denotation, hence no denotation (as "the" implies uniqueness)—unless some additional restriction on the allowable denotation is implicitly assumed. I shall propose to take this idea seriously in my attempt to preserve S4.

## 6. Overlap and Reference Shift

I shall now extend the five-dimensionalist recasting of Chandler's scenario and give my proposal. First, here is my recasting of Chandler's scenario.

Alfred has different worldly parts at $w_0$ and at $w_1$. Call these parts *Alfred-at-*$w_0$ and *Alfred-at-*$w_1$, respectively. Alfred-at-$w_0$ exists at $w_0$ and at no other world, while Alfred-at-$w_1$ exists at $w_1$ and at no other world. When $p_0$ points to an alpha at $w_0$, $p_0$ points to Alfred by pointing to Alfred-at-$w_0$, and thereby succeeds in speaking of Alfred when she says, "The alpha I am pointing to originated from 1-2-3 but could have originated from 4-2-3." She speaks truthfully, for at some world, viz., $w_1$, accessible from $w_0$, Alfred originated from 4-2-3. Alfred originated from 4-2-3 at $w_1$ by having a worldly

[10] This assumes that the modal is just one-dimensional. This is nothing more than a simplifying assumption for the sake of discussion. The modal should probably be considered multi-dimensional. It also assumes that what exists at a world is four-dimensional. Those who regard the number of dimensions of what exists at a world to be less than four could add a modal dimension as an additional dimension but would resist calling the result "five-dimensionalism."

part at $w_1$ which originated from 4-2-3. When $p_1$ points to an alpha at $w_1$, $p_1$ points to Alfred by pointing to Alfred-at-$w_1$, and thereby succeeds in speaking of Alfred when he says, "The alpha I am pointing to originated from 4-2-3 but could have originated from 4-5-3." He speaks truthfully, for at some world, viz., $w_2$, accessible from $w_1$, Alfred originated from 4-5-3. Alfred originated from 4-5-3 at $w_2$ by having a worldly part at $w_2$ which originated from 4-5-3. Here the accessibility relation remains constant; $w_2$ is accessible from $w_1$ in the same sense in which $w_1$ is accessible from $w_0$. The same accessibility relation is in question for both pairs of worlds. So there is no shift from one variety of metaphysical possibility to another variety of metaphysical possibility, as is the case in Murray and Wilson's proposal. One and the same accessibility relation holds between $w_2$ and $w_1$, and between $w_1$ and $w_0$. But at $w_0$ Alfred could not possibly have originated from 4-5-3. Therefore, accessibility is not transitive.

In order to block this five-dimensionalistically recast argument by Chandler, let us return to the elephant parable and modify it a little.

Suppose that instead of one elephant, two elephants—Elphie$_0$ and Elphie$_1$—are standing calmly. Suppose further that Elphie$_0$ and Elphie$_1$ are Siamese twins joined at the belly. When the first blind person touches Elphie$_0$, she touches its trunk but not the trunk of Elphie$_1$, so that when she says, "The animal I am touching is like a snake," she is speaking of Elphie$_0$, and not Elphie$_1$. When the second blind person touches the elephant belly, he is touching the common part of Elphie$_0$ and Elphie$_1$, so that when he says, "The animal I am touching is like a wall," he is not speaking of Elphie$_0$ to the exclusion of Elphie$_1$, or of Elphie$_1$ to the exclusion of Elphie$_0$. The definite description "the animal I am touching" in his mouth fails to denote a unique animal.[11] If it denotes at all, it denotes ambiguously. Or we may say that it denotes a unique animal equivocally. It is open to two different but equally good interpretations, according to each of which it denotes a unique animal. When the story is told in such a way that Elphie$_0$ is introduced first and then Elphie$_1$ is mentioned only later, it is natural and tempting—because of the overlap—to see the second blind person as touching the slightly more familiar elephant, viz., Elphie$_0$, so that we end up being drawn to the interpretation of his definite description "the animal I am touching" as denoting Elphie$_0$.

With this picture of Siamese twin elephants clear in mind, let us return to Alfred. What transpires in the situation concerning Alfred is analogous to that concerning the Siamese twin elephants. When we say that $p_0$ truthfully says of Alfred that it could have originated from 4-2-3, we refer by our use of the name "Alfred" to a certain modally extended alpha which originated from 1-2-3 at $w_0$. So far so good. But just as the second blind person's term "the animal I am touching" is equivocal, our term "Alfred" is equivocal as we use it to describe $p_1$.[12] When we describe $p_1$ as saying of Alfred that it could have originated from 4-5-3, our use of "Alfred" is open to two

[11] Assuming that the two Siamese elephants together as a whole are not one animal.

[12] "Alfred" is also equivocal as we use it to describe what $p_0$ says, as will become clear once the entire picture is in view. But dialectically we need here to focus on "Alfred" as we use it to describe what $p_1$ says.

different and equally good ways of understanding. When understood one way, it refers to one alpha, and when understood the other way, it refers to another, different alpha.

At $w_0$ Alfred could have originated from 4-2-3. So, we say, there is a world $w_1$ accessible from $w_0$ such that at $w_1$ Alfred originated from 4-2-3. When we say this, the term "Alfred," as it occurs in our clause "at $w_1$ Alfred originated from 4-2-3," is quite naturally understood to be coreferential with "Alfred" as it occurs in the preceding sentence "At $w_0$ Alfred could have originated from 4-2-3." This corresponds to the interpretation of "the animal I am touching" as uttered by the second blind person according to which it denotes the same animal as it does when uttered by the first blind person, that is, it denotes $\text{Elphie}_0$.

But after noting that at $w_1$ Alfred originated from 4-2-3, we also say that at $w_1$ Alfred could have originated from 4-5-3, and therefore at some world $w_2$, accessible from $w_1$, Alfred originated from 4-5-3. When we say this, it is more natural and charitable to understand the term "Alfred" occurring in the clause "at … $w_2$ … Alfred originated from 4-5-3" as referring to another, different alpha, of which it is true to say that it has a worldly part that originated from 4-5-3. This corresponds to the interpretation of "the animal I am touching," as uttered by the second blind person, according to which it denotes $\text{Elphie}_1$.[13] We are using the name "Alfred" to refer to one alpha and then to refer to another alpha. This is the shift that saves transitivity of accessibility.

Natural and charitable as they are, these shifty readings of our use of the name "Alfred" are not inevitable, and one may insist on the non-shifty reading according to which all of our uses of "Alfred" refer to just one object. Is such a reading compatible with five-dimensionalism? Does it preserve transitivity of accessibility?

It is certainly compatible with five-dimensionalism, for one may say with legitimacy that the one object "Alfred" refers to in all of our uses is a five-dimensionally spread-out object whose parts include the respective worldly parts in question at the worlds, $w_0$, $w_1$, and $w_2$. If such an object is an alpha, then accessibility is not transitive, as shown by Chandler's argument. So in order to resist the argument, we need to say that such an object is not an alpha. And here again the elephant analogy helps us. We can say that such an object is not an alpha any more than the entire Siamese-twin elephants as a whole are an elephant. We have two elephants, not one. Likewise in the case of "Alfred," we have two alphas,[14] not one. The whole object consisting of the two elephants is not itself an elephant, but it is still something. But since it is not an elephant, the condition of origin applicable to elephants need not apply to it. Similarly, the five-dimensionally spread-out object having the worldly parts at $w_0$, $w_1$, and $w_2$ is something. But since it is not an alpha, the condition of origin applicable to alphas need not apply to it. In other words, if we insist on using "Alfred" to refer to a

[13] I have not supplied a detailed narrative about $\text{Elphie}_1$ which would make the correspondence more vivid, but it should not be difficult to do so.

[14] *At least* two alphas. We really have more than two alphas, but what is important is that we have more than one, that is, we do not have a uniquely determined alpha.

single five-dimensionally spread-out object whose parts include the relevant worldly parts at $w_0$, $w_1$, and $w_2$, then what we refer to by "Alfred" is not an alpha. So, Chandler's argument will lose its foothold.

Suppose that we disambiguate the name "Alfred" and call the alpha (the original alpha) we refer to when considering $w_0$ *Alfred*$_0$, and the alpha we refer to when considering $w_1$ *Alfred*$_1$. Then Chandler's argument may be formulated as follows:

$P_a$: Alfred$_0$ originated from 4-5-3.
$P_b$: Alfred$_1$ originated from 4-5-3.
1". $P_a$ is not true at any world accessible from $w_0$.
2". $P_a$ is not possibly true at $w_0$.
3". $P_b$ is true at $w_2$, and $w_2$ is accessible from $w_1$, which is accessible from $w_0$.
4". $P_a$ is possibly possibly true at $w_0$.
5". Some proposition, viz., $P_a$, is not possibly true but possibly possibly true, at $w_0$.
6". Accessibility is not transitive, i.e., Axiom 4 is false.

3" does not yield 4", which is needed, along with 2", to yield 5". 3" does yield "$P_b$ is possibly possibly true at $w_0$," but this does not help us reach 5", for 2" concerns not $P_b$ but $P_a$. If we replaced 2" with "$P_b$ is not possibly true at $w_0$," then the argument's validity would be restored. But nothing in Chandler's scenario shows that $P_b$ is false at every world accessible from $w_0$, so such a replacement is unsupported.

## 7. Impossible Worlds

Still, it is true to say that at $w_0$ Alfred$_0$ could not have originated from 4-5-3, i.e., at $w_0$ it is impossible that Alfred$_0$ originated from 4-5-3. And according to Chandler, Alfred$_0$ indeed exists and originated from 4-5-3 at $w_2$. So assuming that $w_0$ is the actual world, $w_2$ is an impossible world, according to Chandler. So, Chandler should be happy to place Alfred$_0$ at an impossible world, and Salmon quite explicitly does so and emphasizes the impossibility of the world $w_2$. How do I respond to this stance on an impossible world by Chandler-Salmon?

One way to respond is to downplay the significance of impossible worlds. This is the way of Murray and Wilson.[15] But we need not follow them. We can perfectly well go along with taking impossible worlds seriously. Assuming that $w_0$ is the actual world, according to Chandler-Salmon, $w_2$ is an impossible world, for $w_2$ is not accessible from $w_0$. Although I do not want to downplay the significance of impossible worlds, I think it is a mistake to regard $w_2$ as inaccessible from $w_0$. At $w_2$, an alpha indeed originated form 4-5-3 but it is not Alfred$_0$. Alfred$_0$ does not exist at $w_2$. The alpha which exists and originated from 4-5-3 at $w_2$ is Alfred$_1$.

At $w_0$ Alfred$_0$ originated from 1-2-3 and could have originated from 4-2-3. So at some accessible world, $w_1$, Alfred$_0$ originated from 4-2-3. A different alpha sneaks into the scene at this point, namely, Alfred$_1$. At $w_1$, Alfred$_1$, like Alfred$_0$, originated from 4-2-3 and could have originated from 1-2-3, but unlike Alfred$_0$, it also could have originated from 4-5-3. Conflating the two alphas is at the core of Chandler's error, as we

[15] Murray and Wilson say that Salmon's mention of an impossible world is a distraction.

have observed. But there is a third alpha, which just sneaked into the picture as we spoke of possible origination from 4-5-3 at $w_1$. At $w_2$, which is accessible from $w_1$, $\text{Alfred}_1$ originated from 4-5-3. That is, $\text{Alfred}_1$'s $w_2$-stage originated from 4-5-3. Just as $\text{Alfred}_0$ overlaps $\text{Alfred}_1$ at $w_1$, $\text{Alfred}_1$ overlaps another, third alpha at $w_2$. Interestingly, Chandler, who does not show awareness of the second alpha ($\text{Alfred}_1$), explicitly mentions what is effectively this third alpha; he calls it *Bernard*.[16] Does the introduction of Bernard force an impossible world upon us? No, it does not. Let us see why not.

At $w_0$ it is possible that $\text{Alfred}_0$ never exists but parts 4-5-3 do. So at some accessible world, call it $w_3$, $\text{Alfred}_0$ does not exist but 4-5-3 do. Suppose further that at $w_3$ an alpha originated from 4-5-3. This alpha is Bernard. Since $w_3$ is accessible from $w_0$, and 4-5-3 and 1-2-3 have only 3 in common, Bernard is not $\text{Alfred}_0$. $\text{Alfred}_0$ has no worldly part at $w_3$ or at $w_2$. Bernard has a worldly part at $w_1$, which is the $w_1$-stage of an alpha originating from 4-2-3. But of course, that $w_1$-stage is the $w_1$-stage of $\text{Alfred}_1$ and of $\text{Alfred}_0$ as well; Bernard overlaps $\text{Alfred}_1$ and $\text{Alfred}_0$ at $w_1$. Bernard overlaps $\text{Alfred}_1$, but not $\text{Alfred}_0$, at $w_2$.

I simply identify $w_3$ with $w_2$. Nothing forces this identification; it is possible to maintain that $\text{Alfred}_1$ has no worldly part at $w_3$. But at the same time, nothing forbids the identification, and the identification simplifies the picture. When someone says, having in mind the common $w_1$-stage in question, "Alfred could have originated from 4-5-3," by the name "Alfred" she either refers to $\text{Alfred}_1$ or Bernard and says something true, or else refers to $\text{Alfred}_0$ and says something false. This is how I block Chandler's argument.

Still, I agree with Chandler-Salmon that $\text{Alfred}_0$ exists and originated from 4-5-3 at some impossible world. It is just that I deny that $w_2$ ($=w_3$) is such a world. Let $w_{20}$ be such a world. For Chandler-Salmon's purposes, $w_{20}$ is $w_2$. Chandler-Salmon's argument is in effect that since $w_{20}$ ($= w_2$) is accessible from $w_1$, $w_1$ is accessible from $w_0$, and $w_{20}$ ($= w_2$) is inaccessible from $w_0$, accessibility is not transitive. I agree that $w_1$ is accessible from $w_0$ and that $w_{20}$ is inaccessible from $w_0$, but I deny that $w_{20}$ is accessible from $w_1$. To think that $w_{20}$ is accessible from $w_1$ is to conflate $\text{Alfred}_0$ with $\text{Alfred}_1$.[17]

It is important not to forget that even though an alpha originated from 4-5-3 at $w_2$, it is a different alpha from $\text{Alfred}_0$; it is $\text{Alfred}_1$. $\text{Alfred}_0$ originated from 4-5-3 at $w_{20}$ instead. The two worlds, $w_2$ and $w_{20}$, may well be qualitatively indistinguishable. Even so, they are distinguishable with respect to the identity of the alpha originating from 4-5-3; they are two distinct worlds.[18] Moreover, they are differently related to

[16] Chandler credits Robert Stalnaker for suggesting the Bernard example.

[17] Or worse, with Bernard.

[18] This raises an interesting issue of haecceitism, which is the claim that for any possible worlds $w$ and $w'$, if $w$ and $w'$ are qualitatively indistinguishable, then they are indistinguishable *simpliciter* (see Lewis 1986). In particular, two possible worlds agreeing in all matters qualitative agree in all matters *de re*. The pair of worlds $\{w_2, w_{20}\}$ might look like a counterexample to haecceitism in this sense. But they are not; for a genuine counterexample needs to be a pair of possible worlds, and $w_{20}$ is not a possible world (assuming that $w_0$ is the actual world).

other worlds: $w_2$ is accessible from $w_0$ and from $w_1$, but $w_{20}$ is not accessible from $w_0$ or from $w_1$.

My position is that no impossible world is accessible from a possible world. Assuming that $w_0$ is the actual world, "possible world" means "world accessible from $w_0$" and "impossible world" means "world inaccessible from $w_0$." Moreover, I am assuming with Chandler-Salmon that $w_1$ is accessible from $w_0$ and $w_{20}$ is inaccessible from $w_0$, and claiming that $w_{20}$ is inaccessible from $w_1$. This claim of mine then amounts to the claim that even though $\text{Alfred}_0$ could possibly have originated from 4-2-3, if it had so originated, it would not be possible for $\text{Alfred}_0$ to have originated from 4-5-3. But in making this claim, am I not flouting origin essentialism about alphas? Am I not simply denying the principle that for any alpha, if it originated from x-y-z, then it is possible for it to have originated from x-v-z, where $v \neq y$?

Someone might respond on my behalf by making a distinction between two principles of origin essentialism and affirming one of them, while denying the other. The two principles to be distinguished are as follows:

(E1) For any alpha, if it originated from x-y-z, then it is possible for it to have originated from three parts two of which are x, y, or z.

(E2) For any alpha, if it could have originated from x-y-z, then it could have been possible to have originated from three parts two of which are x, y, or z.

My hypothetical spokesperson might affirm (E1) and deny (E2).[19] In the possible-worlds framework, this asymmetric treatment of the two principles amounts to privileging the actual world. If an alpha originated from x-y-z at the actual world, it originated from x-v-z at some world accessible from the actual world; but if an alpha originated from x-y-z at a non-actual world *w*, there might or might not be a world accessible from *w* at which it originated from x-v-z. Such privileging of actuality is in concert with the fact, of which Saul Kripke famously reminded us,[20] that we standardly discuss non-actual possibilities by starting with actually existing objects and stipulatively considering non-actual possibilities concerning *them*.

Suppose that $w_0$ is the actual world and that at $w_0$ you point to an alpha which originated from 1-2-3 and say truthfully, "This is Alfred and Alfred is an alpha that could have originated from 4-2-3." Suppose also that at $w_1$ someone like you points to an alpha which originated from 4-2-3 and say truthfully, "This is Alfred and Alfred is an alpha that could have originated from 4-5-3." These statements are both true, for you refer to $\text{Alfred}_0$ by "Alfred" at $w_0$ and the person in question at $w_1$ refers to $\text{Alfred}_1$ by "Alfred" at $w_1$. This generalizes to yield the claim that *standardly* no two people attaching a name to an alpha by ostension at two different worlds are speaking of the same five-dimensional alpha; the alphas they are speaking of may well overlap each other extensively, but their five-dimensional extents are not exactly the same. (If, instead of alphas, we have objects which are governed by vague origin conditions for

[19] (E1) is a metaphysical claim of origin essentialism. I am assuming that all parties involved in the discussion of the Chandler-Salmon argument regard it as a necessary truth (at least for the sake of argument), hence a legitimate starting point in modal reasoning.

[20] Kripke 1980: 44.

numerical identity, then we may be able to avoid this and ensure the exact sameness of the extents by having an account of vagueness concerning the number of parts essential for origin such that the non-overlapping worldly parts of the objects lie well within the penumbra allowed by the vagueness.)

This, of course, does not mean that when speaking at $w_0$, you cannot speak of a particular alpha as having originated from 4-5-3 at some world other than $w_0$. You may perfectly well point to an alpha in front of you and say, "This is Alfred and Alfred is an alpha which could have originated from 4-2-3, and if it had originated from 4-2-3, then it would have been possible for it to have originated from 4-5-3." You may perfectly well be speaking of $\text{Alfred}_0$ throughout your remark; the subject matter of your speech may well remain to be uniquely $\text{Alfred}_0$. But if it does so remain, then the second conjunct of your statement is false; it is false that if $\text{Alfred}_0$ had originated from 4-2-3, then it would have been possible for $\text{Alfred}_0$ to have originated from 4-5-3. At $w_1$ $\text{Alfred}_0$ has a worldly part which originated from 4-2-3, and it is also a worldly part of $\text{Alfred}_1$. And it is $\text{Alfred}_1$, not $\text{Alfred}_0$, that could at $w_1$ possibly have originated from 4-5-3.

This seems to work as a defense of my proposal. Should I accept (E1), reject (E2), and let my hypothetical spokesperson speak on my behalf then? I am afraid not. Privileging of actuality is an interesting idea and the spirit in which it is proposed will prove to be productive, as we shall see shortly. But if we wish to preserve standard quantified modal logic (SQML)—and I do, as I endeavor to defend Axiom 4, and also Axiom 5 later—then we cannot reject (E2) while accepting (E1); for in SQML, (E1) entails (E2). Let "v" be a restricted variable ranging just over alphas, and let "Fv" and "Gv" mean "v originated from x-y-z" and "v originated from three parts two of which are *x*, *y*, or *z*," respectively. Then (E2) is derived from (E1), as follows:

| | |
|---|---|
| $\forall v(Fv \rightarrow \Diamond Gv)$ | (E1) |
| $\Box\forall v(Fv \rightarrow \Diamond Gv)$ | by the rule of necessitation |
| $\forall v\Box(Fv \rightarrow \Diamond Gv)$ | by the Converse Barcan Formula |
| $\forall v(\Diamond Fv \rightarrow \Diamond\Diamond Gv)$ | (E2) by the valid schema $\Box(\varphi \rightarrow \psi) \rightarrow (\Diamond\varphi \rightarrow \Diamond\psi)$[21] |

This means that within SQML we cannot privilege actuality by means of a distinction between (E1) and (E2), or anything that entails the distinction. This is where the crucial idea of reference shift proves useful.

Some might cast doubt on the first step in the above argument, from (E1) to its necessitation, on the ground that (E1) is not a theorem of SQML. Such a doubt is allayed when we note that the argument starts with (E1) not because (E1) is a theorem of SQML (it certainly is not) but because, as indicated in footnote 19, all parties accept it as a metaphysical claim which does not just happen to be true but is necessarily true. Since Chandler takes himself to be arguing in effect against SQML as he argues under this assumption, it is permissible for me to make the same assumption.

Reference shifts from world to world as we consider whether what Chandler calls "Alfred" had or could have had a certain origin. Along with this reference shift, we

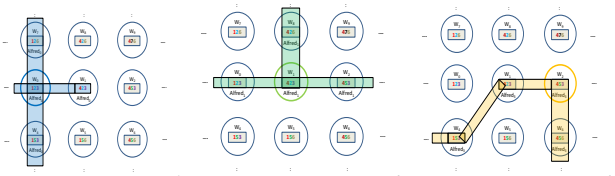
[21] I owe this proof to Alessandro Torza.

should shift the way we apply the idea of essentiality of origin. In fact, without the latter shift, the reference shift alone would be rather pointless for my purposes. When we consider Chandler's alpha at $w_0$ and say, "At $w_0$ $\text{Alfred}_0$ could have originated from 4-2-3," we are applying the principle of origin essentialism correctly, but if we moved our consideration to $w_1$ and said, "At $w_1$ $\text{Alfred}_0$ could have originated from 4-5-3," we would not be applying the principle of origin essentialism correctly. When we shift our talk to $w_1$ with Chandler, we shift our reference of the name "Alfred." This is my claim of reference shift. Resorting to such reference shift would be moot unless application of the principle of origin essentialism is also adjusted so that the new referent is the subject matter when we say that *it* could have originated from 4-5-3.

On my proposal of reference shift, we refer to different alphas when considering different worlds in accordance with Chandler's scenario. The correct way to apply the principle is to say that at $w_0$ $\text{Alfred}_0$ could have originated from 4-2-3, that at $w_1$ $\text{Alfred}_0$ could have originated from 1-2-3, that at $w_1$ $\text{Alfred}_1$ could have originated from 4-5-3 or from 1-2-3, that at $w_1$ Bernard could have originated from 4-5-3, and that at $w_2$ Bernard could have originated from 4-2-3. It is an incorrect application of the principle to say that at $w_1$ $\text{Alfred}_0$ could have originated from 4-5-3, or that at $w_1$ Bernard could have originated from 1-2-3.

The principle of origin essentialism for alphas allows one-third deviation in origin away from, or toward, the *modal center* of the alpha in question, but in no other direction. The (modal) center of $\text{Alfred}_0$ is its $w_0$-stage, the center of $\text{Alfred}_1$ is its $w_1$-stage, and the center of Bernard is its $w_2$-stage. Since $\text{Alfred}_0$'s center is its $w_0$-stage, which originated from 1-2-3, at $w_0$ $\text{Alfred}_0$ could have originated from, for example, 4-2-3, 7-2-3, 1-10-3, 1-11-3, 1-2-13, or 1-2-14 (away from the center), and at $w_1$ $\text{Alfred}_0$ could have originated from 1-2-3 (toward the center). But at $w_0$ $\text{Alfred}_0$ could not have originated from, for example, 1-5-6, 4-2-6, or 4-5-3 (too far away from the center). At $w_1$, where $\text{Alfred}_0$ is not centered, $\text{Alfred}_0$ could have originated from, for example, 1-2-6 or 1-5-3 (to the center then away from it). Thus, assuming that $\text{Alfred}_0$ originated from 1-2-6 at $w_7$, $w_7$ is accessible from $w_1$, even though at $w_1$ $\text{Alfred}_0$ originated from matter which is two-thirds different from the matter from which it originated at $w_7$.

It is natural that the appearance to the contrary is created and we are tempted to answer "No" when we ask ourselves the question, "At $w_1$ could $\text{Alfred}_0$ have originated from 1-2-6?" This is because when we fix our attention on possibilities at $w_1$, we are naturally led to assume that we are speaking of an alpha centered at $w_1$. But since we are in fact speaking of $\text{Alfred}_0$ and $\text{Alfred}_0$ is centered at $w_0$, not at $w_1$, the answer we are tempted to give is the wrong answer.



$\text{Alfred}_0$ $\text{Alfred}_1$ Bernard

In general, assuming that $\text{Alfred}_k$'s center is its $w_k$-stage, which originated from x-y-z, at $w_k$ $\text{Alfred}_k$ could have originated from matter consisting of at least two of x, y, and z (away from the center) but not from matter consisting of no more than one of x, y, and z; and at $w_{k+1}$—where $\text{Alfred}_k$'s origin differs from x-y-z by exactly one part, say, x-y-v—$\text{Alfred}_k$ could have originated from x-y-z (toward the center) or from u-y-z or x-u-z, where u is neither x nor y (to the center then away from it), but not from more different matter than these. Thus the principle of origin essentialism should say:

(OE) For every alpha, it is necessary that it originated from matter that is at most one-third different from the matter its center originated from (and it is possible that it originated from matter that is only one-third different from the matter its center originated from).

Alphas are artificially well-behaved objects. Ordinary objects are much more complicated. Their origin has many more parts than three, and the principle of origin essentialism for them are much harder to formulate. Still, the basic idea applies to them just as well. Each object is a five-dimensional object with a center, which consists of many world-stages instead of just one. A clear line (like "one-third" for alphas) cannot be drawn, and the issue of vagueness needs to be faced squarely. But these are mere complications, rather than fundamentally different considerations that change the shape of the discussion.

## 8. Euclidean-ness

Chandler's thought experiment is easily adaptable to produce an argument against the characteristic axiom of S5, viz., Axiom 5: $\Diamond A \rightarrow \Box\Diamond A$. This axiom requires that the accessibility relation be euclidean, that is, if two worlds are accessible from a common world, then they are mutually accessible. In the middle diagram above, $\text{Alfred}_1$ originated from 1-2-3 at $w_0$, from 4-2-6 at $w_8$, and from 4-2-3 at $w_1$. So, $w_0$ and $w_8$ are both accessible from $w_1$, but $\text{Alfred}_1$'s origin at $w_0$ and $\text{Alfred}_1$'s origin at $w_8$ have only one common part; so neither of $w_0$ and $w_8$ is accessible from the other, someone might say. Therefore, it might be concluded, accessibility is not euclidean, hence Axiom 5 is false.[22]

This objection against Axiom 5 is answerable by resorting to the idea of the center of a modally extended object and using the principle of origin essentialism (OE). $\text{Alfred}_1$ is centered at $w_1$, and $\text{Alfred}_1$'s $w_0$-stage and $w_8$-stage both originated from matter having two common parts with the matter from which $\text{Alfred}_1$'s $w_1$-stage originated. Thus, according to (OE), $\text{Alfred}_1$ at $w_0$ and $\text{Alfred}_1$ at $w_8$ do not flout origin essentialism. We are assuming that nothing else stands in the way of mutual accessibility between $w_0$ and $w_8$. Therefore, $w_0$ and $w_8$ are accessible from each other.

[22] Since $w_0$ is the (presumed) actual world, this consideration gives us another reason why distinguishing (E1) from (E2) does not help. I owe this point to Axel Barceló.

## 9. Obstinate Essentialism

Recall that obstinate essentialism says that any object that originated from matter consisting of certain parts necessarily originated from matter consisting of exactly those parts. So, according to obstinate essentialism, $\text{Alfred}_0$ necessarily originated from 1-2-3, that is, for any possible world $w$, an alpha originating from any matter other than 1-2-3 at $w$, is not $\text{Alfred}_0$. Some philosophers may find this position appealing. My proposal of reference shift explains the apparent appeal of obstinate essentialism without endorsing it.

Chandler's scenario starts with an alpha originating from 1-2-3 at $w_0$; that alpha is $\text{Alfred}_0$. According to my proposal of reference shift, when we consider an alpha at $w_1$ originating from 4-2-3, we are naturally and most likely referring to $\text{Alfred}_1$, not $\text{Alfred}_0$; that is why it is natural and most likely true to say then, "At $w_1$ Alfred could have originated from 4-5-3." When we consider the relevant alpha at yet another relevant world, $w_2$, we are naturally and most likely referring to Bernard, not $\text{Alfred}_0$ or $\text{Alfred}_1$. (The relevant alpha at a relevant world is an alpha that is different from the previous alpha in origin by one third and existing at a world minimally different from the previous world.) These are all different alphas, hence the impression is created that no alpha survives the minimal amount of change in origin across worlds.

At the same time, what is overlooked by the obstinate essentialist is that since $\text{Alfred}_0$ and $\text{Alfred}_1$ overlap at $w_1$, having the common $w_1$-stage, which originated from 4-2-3, we (at $w_0$) can refer to $\text{Alfred}_0$ rather than $\text{Alfred}_1$ by pointing (with the mind's finger) to that $w_1$-stage and considering it in conjunction with considering what exists at $w_0$, in particular, the $w_0$-stage of an alpha. When we do so, we speak truthfully by saying, "Alfred originated from 1-2-3 at $w_0$ but originated from 4-2-3 at $w_1$, so it could have originated from matter slightly different from the matter it actually originated from." By contrast, if we point to the same $w_1$-stage and consider it not in conjunction with considering the $w_0$-stage of an alpha existing at $w_0$, but afresh as the starting point of discussing origin essentialism, then we may well be able to speak truthfully by saying, "Alfred originated from 4-2-3 at $w_1$, so it could have originated from 4-5-3," referring to $\text{Alfred}_1$ by "Alfred."[23]

## 10. Alternative Modal Space

The thesis I defend says that the accessibility relation underlying metaphysical possibility is transitive in modal space. Chandler-Salmon say that it is not transitive. So I say that they get modal space wrong. At the same time, I think that they get something right. Let me explain.

[23] The question which modally extended object we succeed in referring to via a particular worldly stage is not an easy question to answer. When we point to the portion of the body shared by two Siamese twins while intending to refer to one of the twins as opposed to the other, what determines which twin we succeed in referring to? Is our intention alone sufficient for the determination? Or do other contextual factors figure somehow, and if so, what factors and how? The modal case is no easier than the Siamese twin case to settle.

World indexing works well for evaluations of truth-values of ordinary statements about ordinary objects like bicycles, boxes, branches, and bears, but not for evaluations of truth-values of some statements about worlds. Statements of logical properties, like transitivity, concerning the accessibility relation are statements about worlds, as both *relata* of accessibility are worlds. Some relations between worlds may be said to hold or fail to hold only relative to a world. Take closeness, for example; a world may be closer than another world in the sense that the first world resembles a certain given world more than the second world does. In such a case, we may regard the dyadic closer-than relation as holding between two worlds from the point of view of the given world, that is, the dyadic relation as holding between two worlds relative to a world. But accessibility is not like that. A world is accessible from another world (or not), independently of relativization to a given world. What particular accessibility relation is in question in a given discourse is, of course, relative to what is relevant to the discourse; it may be metaphysical accessibility in one discourse, logical accessibility in another discourse, and nomological accessibility in yet another. But once a particular accessibility relation is determined in a given discourse, which world bears that relation to which world is not a matter relative to a world. So whether the particular accessibility relation underlying metaphysical possibility is transitive is not a matter relative to a world. But it is still relative, and this allows room for something that Chandler-Salmon get right.

Whether a given relation has or lacks a given logical property, like transitivity, is determined globally within the entire realm that comprises eligible *relata*. When the given relation is the accessibility relation underlying metaphysical possibility, the relevant realm is a realm comprising worlds. If for some triple of worlds $w$, $w'$, and $w''$ in the realm, accessibility holds between $w$ and $w'$, and between $w'$ and $w''$, but not between $w$ and $w''$, then it is correct to say relative to the realm that accessibility is not transitive; otherwise, it is correct to say relative to the realm that it is transitive. I think that since there is no such triple of worlds relative to a certain realm we (implicitly) have in mind when discussing Chandler-Salmon's challenge to S4, accessibility is transitive relative to the realm. The realm in question is the realm comprising all the worlds standardly considered when *local* metaphysical possibility is in question, viz., *local modal space,* and this is the exact sense in which we say that accessibility is transitive *in modal space*.[24]

Now, when Chandler-Salmon maintain that accessibility is not transitive, I understand them as maintaining that accessibility is not transitive in modal space. Since, in my opinion, accessibility is transitive in modal space, I say that Chandler-Salmon are wrong. But, in their arguments Chandler-Salmon do successfully describe some realm. They simply mistake it for modal space and end up maintaining a falsity. But

[24] The adjective "local" modifies the noun phrase "modal space" in a way parallel to the way in which "actual" modifies "world." If the actual world is the default world of our discourse, the local modal space is the default modal space of our discourse. Indeed, the parallel is so striking that some of us might even be tempted to use "actual" instead of "local" to modify "modal space."

their description applies correctly to that realm, all the same. And that realm is just like modal space but it is alternative to it. It is a realm in which for some triple of worlds, *w*, *w′*, and *w″*, accessibility holds between *w* and *w′*, and between *w′* and *w″*, but not between *w* and *w″*. This *alternative (non-local) modal space* does indeed comprise a world (*w*) at which Alfred, which actually (at *w′*) originated from 1-2-3, originated from 4-5-3, while remaining an alpha. Chandler-Salmon say that such a world is an impossible world, and I agree. Chandler-Salmon conclude from this—assuming the relevant background information about *w*, *w′*, and *w″*—that accessibility is not transitive, that is, accessibility is not transitive in modal space. I, by contrast, say that Chandler-Salmon change the subject. The impossible world *w* does not belong to modal space, viz., *the local modal space*. It belongs to some alternative modal space, and in that alternative modal space, accessibility fails to be transitive. But that alternative modal space is not the modal space in question—our *local modal space*—but comprises a metaphysically impossible world which robs accessibility of a logical property it in fact has in our local modal space.

## 11. Speculation

There are many modal spaces, and each of them comprises worlds. Our actual world, assumed to be $w_0$, exists in our local modal space—call it $s_0$—and also in other modal spaces. Many other worlds also exist in $s_0$ and in other modal spaces. Every world exists in some modal space, but probably not in every modal space. Let us say that $w_0$ *locally*$_s$ exists in $s_0$ and non-locally$_s$ but *alternatively*$_s$ exists in another modal space. Locality$_s$ and alternativeness$_s$ pertain to modal spaces in a way parallel to the way in which actuality and mere possibility pertain to worlds. The world $w_1$ is accessible from $w_0$ in $s_0$ ($w_1$ is locally$_s$ accessible from $w_0$), but it is not accessible from $w_0$ in some other modal space (it is alternatively$_s$ not accessible from $w_0$). Accessibility is transitive in $s_0$ (accessibility is locally$_s$ transitive), but it is not in some other modal space (it is alternatively$_s$ not transitive). The worlds, $w_0$, $w_1$, and $w_2$, locally$_s$ exist, $w_2$ is locally$_s$ accessible from $w_1$, $w_1$ is locally$_s$ accessible from $w_0$, and $w_2$ is locally$_s$ accessible from $w_0$. The worlds $w_0$ and $w_1$ also alternatively$_s$ exist, along with $w_{20}$, in some alternative modal space, and in that alternative modal space, $w_{20}$ is alternatively$_s$ accessible from $w_1$, $w_1$ is alternatively$_s$ accessible from $w_0$, and $w_{20}$ is alternatively$_s$ not accessible from $w_0$.[25] It is such an alternative modal space that Chandler-Salmon inadvertently end up describing.

[25] When I say this, I do not mean that in that alternative modal space, $w_{20}$ is not but could be accessible from $w_1$, $w_1$ is not but could be accessible from $w_0$, and $w_{20}$ is not but could be inaccessible from $w_0$. Rather, I mean that in that alternative modal space, $w_{20}$ is accessible from $w_1$, $w_1$ is accessible from $w_0$, and $w_{20}$ is inaccessible from $w_0$. The point of the adverb "alternatively$_s$" is to distinguish these predications of accessibility and inaccessibility from predications within local modal space, which in contrast are to be marked by the adverb "locally$_s$." There is more to be said about this and what I call *modal tense*; for the basic idea of modal tense, see Yagisawa 2010, chapter 5.

When we say that accessibility is transitive, what we usually mean is that accessibility is locally$_s$ transitive, just as when we say that the universe is expanding, what we usually mean is that the universe is actually expanding. When we say that since it is possible that Bernard originated from 4-5-3, Bernard originated from 4-5-3 at a possible world, we should mean that Bernard locally$_s$ originated from 4-5-3 at a possible world, just as when we say that Caesar crossed the Rubicon, we should mean that Caesar actually crossed the Rubicon. When, on the other hand, we insist that since it is impossible that Alfred$_0$ originated from 4-5-3, Alfred$_0$ originated from 4-5-3 at an impossible world, we should mean that Alfred$_0$ alternatively$_s$ originated from 4-5-3 in some modal space in which worlds are related in an alternative$_s$ way. I emphasize that when we insist on understanding the statement of the impossibility of Alfred$_0$ having originated from 4-5-3 by locating Alfred$_0$'s having originated from 4-5-3 at an impossible world, our subject matter automatically shifts from the local$_s$ modal space $s_0$ to some alternative$_s$ modal space, call it $s_{20}$, comprising the impossible world $w_{20}$, along with the worlds $w_0$ and $w_1$.

Alfred$_0$'s having an impossible origin 4-5-3 is to be located at a world outside $s_0$, for to speak of Alfred$_0$ having originated from 4-5-3 is to speak of Alfred$_0$ having a worldly part which Alfred$_0$ does not in fact have, i.e., does not have in $s_0$. Such Alfred$_0$ spreads out in some modal space in a way different from the way Alfred$_0$ spreads out in $s_0$. Such Alfred$_0$ spreads out in some alternative modal space, $s_{20}$. The $s_{20}$-stage of Alfred$_0$ has a different shape (spread) from the $s_0$-stage of Alfred$_0$. Alfred$_0$ is extended not just five-dimensionally[26] but also six-dimensionally.[27]

## References

Chandler, H.S. 1976, "Plantinga and the Contingently Possible", *Analysis*, 36, 106-109.

Kripke, S.A. 1980, *Naming and Necessity*, Cambridge (MA): Harvard University Press.

Lewis, D. 1986, *On the Plurality of Worlds*, Oxford: Blackwell.

Murray, A. and Wilson, J. 2012, "Relativized Metaphysical Modality", in Bennett, K. and Zimmerman, D. (eds.), *Oxford Studies in Metaphysics*, Vol. 7, Oxford: Oxford University Press, 189-226.

Salmon, N. 1981, *Reference and Essence*, Princeton, NJ: Princeton University Press.

[26] Thus the title of this paper should be taken with a grain of salt. In fact, the number of dimensions of Alfred$_0$'s spread does not stop at six. Alfred$_0$ is not an $n$-dimensional object, for any finite $n$. But this is a further speculative idea to be explored on another occasion.

[27] Shorter versions of this paper were presented at the following conferences in 2014: Veritas Philosophy Conference, Yonsei University, Seoul, South Korea, June 8; Intensional/Hyperintensional, National Autonomous University of Mexico, Mexico City, September 18; Modal Metaphysics: Issues on the (Im)Possible II, Institute of Philosophy of Slovak Academy of Sciences, Bratislava, Slovakia, October 15. I thank the audiences for useful discussion. I am particularly grateful to Axel Barceló and Alessandro Torza for helpful challenges. I also thank the two anonymous referees for *Argumenta*.

Salmon, N. 1984, "Impossible Worlds", *Analysis*, 44, 114-17.

Salmon, N. 1989, "The Logic of What Might Have Been", *The Philosophical Review*, 98, 3-34.

Yagisawa, T. 2010, *Worlds and Individuals, Possible and Otherwise*, Oxford: Oxford University Press.

# World Stories and Maximality

*Vittorio Morato*

*University of Padua*

*Abstract*

According to many actualist conceptions of modality, talk about possible worlds should be reduced to talk about world stories. Intuitively, a world story is a complete description of how things could be. In this paper, I will claim that the world story approach not only suffers from the well-known, expressive problem of representing the thesis of the possible existence of non-actual objects, but it has troubles in representing, in an actualistically acceptable way, the apparently more tractable thesis of the possible non-existence of actual objects. To solve this problem, I will propose a refinement of the approach by the introduction of a novel notion of maximality, local maximality.

*Keywords*: modality, world stories, actualism, possibilism, maximality.

## 1. Introduction

According to many *actualist* conceptions of modality, talk about possible worlds should be reduced to talk about *world stories*.[1] Intuitively, a world story is a *complete description of how things could be*. Formally, a world story is a certain set of (actually existing) propositions that is *consistent* and *maximal*.

Consistency and maximality for sets of propositions are usually defined in the following way:

**Consistency:** a set Γ of propositions is consistent if and only if it is possible that all members of Γ be true together.[2]

[1] "Actualism" is often characterized as the thesis that the only objects that exist are those *actually* existing or, in other terms, that to exist is to be actual (Menzel 2016, Divers 2002). World stories were first introduced in Adams 1974 and further discussed, in a much more detailed way, in Adams 1981.

[2] This way of defining consistency of a set of propositions presupposes a certain notion of possibility as primitive. There is at least one way to define consistency in a *non-modal* way, but the definition would work for (uninterpreted) sentences, not propositions: a set Γ of sentences is consistent iff it is satisfiable, where a set Γ of sentences is satisfiable iff there is a propositional interpretation that makes all members of Γ true.

*Argumenta* 2, 2 (2017): 263-279
ISSN 2465-2334


DOI 10.23811/45.arg2017.mor

**Maximality:** a set Γ of propositions is maximal if and only if, for every pair of mutually inconsistent propositions $p_1$ and $p_2$, either $\Gamma \in p_1$ or $\Gamma \in p_2$.

The paradigmatic pair of mutual inconsistent propositions is the one formed by a proposition $p$ and its sentential negation $\neg p$. If one believes that the only kind of mutual inconsistency between pairs of propositions is the one expressed by sentential negation, the definition of maximality above amounts to the following: a set Γ of propositions is maximal if and only if, for every proposition $p$, either $p \in \Gamma$ or $\neg p \in \Gamma$. The idea is that maximality accounts for the fact that a world story is a *complete* description of how (all) things could be—a possible world is a "*total* history" as Kripke writes[3]—while consistency accounts for the fact that a world story is a description of a *possible* way things could be and whatever is true within (even a non-total) possibility should be compatible with whatever else is true within that very same possibility.

A *Russellian conception of propositions* is often associated with this view. According to such a conception, propositions are *structured* entities and there are some propositions, called *singular propositions*, that are about a certain object by having that object as a *direct* constituent of the proposition itself. The view that propositions are structured entities is sometimes called *structuralism*, the view that propositions may have objects as their direct constituents is sometimes called *objectualism*. The Russellian conception of propositions is then the result of combining objectualism and structuralism about propositions. World stories are then typically conceived as maximal and consistent sets of Russellian propositions. Notice that, from what I have claimed so far, world stories could be taken as maximal and consistent sets of both singular and non-singular, i.e., general, Russellian propositions.

The Russellian conception of propositions *plus* actualism implies that the objectual components of the propositions belonging to a world story will be actual objects. If the world stories theorist believes that not every actual object exists necessarily, or in a stronger way, that every actual object exists contingently, then she will also be committed to the view that some of her propositions (some or all of those having actual objects as direct constituents) will be contingent existents and thus that her world stories will be contingent existents too.[4]

The association between the world stories approach and the Russellian conception of propositions is probably the most philosophically sensible one to have, but it is not a forced one. The world stories approach is compatible with non-objectual conceptions of propositions. For example, it is compatible with a Fregean conception of propositions, according to which propositions are structured entities whose immediate components are intensional entities of one sort or another, and not individuals.

[3] Cf. Kripke 1980: 15.

[4] On the contingency of structural and objectual propositions, see Fine 1980: 161. For a world story to be contingent it is enough that at least one of its members be contingent; for a proposition to be contingent it is enough that at least one of its components be contingent. One might, of course, deny that from the contingency of the constituents of a structure $S$ something follows about the modal status of $S$. In order to do that, however, one should deny that $S$ ontologically depends on its constituents (or deny that the relation of ontological dependence between components $a_1, \ldots, a_n$ and a structure $S$ does not transfer the modal status of $a_1, \ldots, a_n$ to $S$).

In this paper, I will claim that the world stories approach not only suffers from the well-known, expressive problem of representing, in an actualistically acceptable way, the thesis of the *possible existence of non-actual objects*, but it has troubles in representing the apparently more tractable thesis of the *possible non-existence of actual objects*. If an actualist approach has problems in representing both theses, then it could only be associated with a form of *necessitism*, namely the claim that, necessarily, every actual object necessarily exists (Williamson 2013). However, many actualists, and many world stories theorists among them, are not necessitists.

Unfortunately, I do not think that the world stories approach has the resources to represent, in an actualistically acceptable way, the thesis of the possible existence of non-actual objects, but there is at least some hope to make the approach able to represent the thesis of the possible non-existence of actual objects. I will propose a solution to this latter problem based on a refined conception of maximality that I will call "local maximality".

## 2. Two Ways of Building World Stories

In order to give a reductive analysis of possible worlds, two things need to be done: one has to specify what kinds of entities should go proxy for possible worlds and one has to explain what it is for something to be true "according to" (if not even "into") such entities. The fundamental move in possible worlds semantics is the relativization of truth to possible worlds, so, whatever be the kind $X$ of entities to which possible worlds are to be reduced, a corresponding, and plausible, notion of "truth in $X$" needs to be defined.

There are, however, two slightly different ways in which the world stories theorist may define the notion of "truth in/according to a world story", each corresponding to a slightly different conception of what a world story ultimately is.

The first is to define the notion of "true in a world story $S$" in terms of the notion of "belonging to a world story $S$": on this approach, a proposition $\Gamma$ is true in a world story $S$ if and only if $\Gamma$ belongs to the world story $S$. In such a case, the notion of "truth in a world story $S$" is reduced to the notion of "belonging to a world story $S$". This latter notion could then be defined in the following way:

- for any atomic singular proposition $\Phi^n(a_1, \dots, a_n)$ either $\Phi^n(a_1, \dots, a_n) \in S$ or $\neg\Phi^n(a_1, \dots, a_n) \in S$;
- $\neg\Gamma \in S$ if and only if $\Gamma \notin S$;
- $(\Gamma \vee \Delta) \in S$ if and only if $\Gamma \in S$ or $\Delta \in S$;
- $\forall x_1, \dots, x_n\, \Phi^n(x_1, \dots, x_n) \in S$ if and only if, for any actual objects $a_1, \dots, a_n$, $\Phi^n(a_1, \dots, a_n) \in S$;
- $\Box\Gamma \in S$ if and only if, for any world story $S_i$, $\Gamma \in S_i$;
- for any other proposition $\Psi$, $\Psi \notin S$.[5]

The other way consists in recursively defining the notion of "truth in a world story". According to this latter method, however, not everything true in a world story needs to belong to a world story. On this approach, world stories are then

[5] It should be noted that the first three conditions of this definition, while essential to give a working recursive definition of "belonging to a world story S" are really superfluous because they are direct consequences of maximality (the first) and, jointly, of maximality and consistency (the second and the third).

to be conceived as maximal and consistent sets of a *special* class of propositions and a proposition is true in a world story, if it follows from or it belongs to this special class. The special class of propositions by which the notion of truth in a world story is defined is the class of *atomic singular propositions*. As a result, the notions of maximality and consistency need to be tailored for atomic singular propositions:

**Consistency*:** A set Γ of atomic singular propositions is consistent, if and only if it is possible that all members of Γ be true together.

**Maximality*:** A set Γ of atomic singular propositions is maximal* if and only if, for every pair of mutually inconsistent atomic propositions $p_1$ and $p_2$, either $p_1 \in \Gamma$ or $p_2 \in \Gamma$.

Again, if one believes that mutual inconsistency is exclusively expressed by sentential negation, then the definition of maximality* amounts to the following: a set Γ of atomic singular propositions is maximal* if and only if, for every atomic proposition $p$, either $p$ or $\neg p$ belongs to Γ. A world story thus contains either atomic propositions or their negations.

The notion of "truth in a world story S" could be now recursively defined as follows:

- any atomic singular proposition $\Phi^n(a_1, \dots, a_n)$ is true in $S$ if and only if $\Phi^n(a_1, \dots, a_n) \in$ S;
- $\neg\Gamma$ is true in $S$ if and only if Γ is not true in $S$;
- $(\Gamma \vee \Delta)$ is true in $S$ if and only if Γ is true in $S$ or Δ is true in $S$;
- $\forall x_1, \dots, x_n\, \Phi^n(x_1, \dots, x_n)$ is true in $S$ if and only if for any (actual) objects $a_1, \dots, a_n$, $\Phi^n(a_1, \dots, a_n)$ is true in $S$;
- □Γ is true in $S$ if and only if for any world story $S_i$, Γ is true in $S_i$;
- for any other proposition Ψ, Ψ is not true in $S$.

The relation between the two methods of characterising world stories is a typical case of trade off between ontology and ideology: the latter method gives us a much more "austere" version of world stories, but it has to take as primitive the notion of "truth in a world story $S$", the former method reduces the notion of "truth in a world story $S$" to the notion of "belonging to a world story $S$", but it has a much more "inflated" version of world stories. The inflated version of world stories is the one actually used by Robert Adams, while a counterpart of the austere version was the one used by Rudolf Carnap (1947) in *Meaning and Necessity*: state descriptions (roughly, linguistic counterparts of world stories) were conceived as maximal sets of *atomic* sentences of a language $L$.

## 3. The Problem of the Possible Existence of Non-Actuals

Like any other actualist conception of modality, the main problem for the world stories approach is that of representing, in an actualistically acceptable way, the *possible existence of non-actual objects*. As I said in the introduction, it is not my intention, in this paper, to solve this problem for the world stories theorist. Actually, I happen to think that the prospects for a solution are quite dim. Nonetheless, I think it is useful to shortly present the problem.

Take a sentence like:

(1) There could have been a non-actual object;

or something like:

(2) There could have been more objects than there actually are.

Few actualists would be brave enough to deny that such sentences are true.[6] The problem for the actualist is precisely how to accept their truth without accepting also the existence of possible and non-actual objects. For those actualists accepting already the idea that (at least some) actually existing individuals are contingent existents, the possible existence of non-actual objects seems to be a natural thesis to accept.[7]

The view according to which we have possible non-existence of the actuals without the possible existence of the non-actuals corresponds to the view that objects could only fail to exist, while it is not possible for "new" objects to come into existence. Actuality could thus only be different "by subtraction": it is only possible that there exist fewer objects than those that there actually are. This ontological asymmetry would sound quite suspect and it would correspond to an implausible metaphysical view.

The problem for actualists is thus not the mere acceptance of (1) or (2). If one accepts *contingentism* (the view that at least some actual object exists contingently) it would be difficult to deny either of them. The problem is rather that the truth of such sentences seems to be difficult to represent in an *actualistically acceptable way*. There are various ways to define what counts as actualistically acceptable. A criterion of actualistic acceptability may be, for example, that the truths of modal claims should *supervene* on the truth of non-modal claims (where, for the actualist, non-modal truths are the actual truths). From this criterion, it follows that (1) and (2) are not actualistically acceptable, unless one shows that their truth supervenes on the truth of some non-modal claim.

The world stories approach, being an actualist approach, has problems in representing the truth of (1) and (2). Given the conditions above, for the proposition expressed by (1) to be true there should be at least a world story *S* such that the proposition expressed by "there exists a non-actual object" is true in *S*. The non-modal basis over which the truth of (1) should supervene is thus the categorical statement "the proposition that there exists a non-actual object is true in a world story *S*". For this non-modal proposition to be true in *S*, *S* needs to contain a singular proposition that testifies for the general existential proposition that there exists a non-actual object, a singular proposition having an object not satisfying the predicate "being actual". But, given that all the objects there are are actual, no object could satisfy the predicate in question. We then do not have the singular, non-modal, proposition that testifies for the general sentence embedded in (1) and therefore, as the theory stands, our sentence cannot be represented as true in an actualistically acceptable way, simply because it turns out to be false.

[6] So called, "new actualists" (the term comes from Menzel 2016) such as Linsky & Zalta 1994 are an exception: they would rather deny both (1) and (2) and try to explain away the relevant intuitions.

[7] Notice that the actualists could safely speak of the possible existence of non-actual objects without ontologically committing themselves to the existence of possible non-actual objects. The inference from the possible existence of something to the existence of a possible something is granted by the (existentially quantified version of the) Barcan formulas and, typically, actualists do not think that such a formula is valid.

A possible way out for the actualist's may be that of negating that a sentence like (1) needs a singular proposition that makes it true. (1) does not need a testimony, the actualist could claim. A similar view is defended by Kit Fine, according to which the singularity needed to make such a sentence true is "spurious":

> For the actualist […] there can be no instance in virtue of which the sentence is true. The sentence states an irreducible general possibility, and no matter how well the individual is described, he can have no specific identity (Fine 1977: 117).

Many forms of actualism could be characterized as a strive to substantiate Fine's quote (and in particular to substantiate the view that truths about merely possible existents are irreducibly general). Notice, however, that if we take the route of *irreducible general possibilities*, then we should abandon the criterion of supervenience of the modal over the non-modal as a way to represent, in an actualistically acceptable way, alleged truths about possible non-existents. If modality is primitive, modal truths do not supervene on non-modal ones.

## 4. The Problem of the Possible Non-Existence of Actuals

Many energies have been spent on the problem of representing the possible existence of non-actuals in an actualistically acceptable way. Fairly enough, such a problem has been taken to be as *the* problem for actualist approaches to modality. However, few have noticed that the world story approach and, probably, many other actualist approaches, also have problems in representing the seemingly more tractable thesis of the *possible non-existence of an actual object*, namely the problem of representing, in an actualistically acceptable way, a situation in which, for example, I (undoubtedly an actual object) do not exist. The rest of the paper will be devoted to show why the world stories approach has this problem and how it could be solved.

World stories (being sets of propositions) are *representational entities*, they represent things as being in a certain way.[8] In order to see whether the world story theorist is able to represent the possible non-existence of an actual object we should reflect, at least for a moment, on what does it mean for a representation to be a representation of the non-existence of an actual object.

My possible non-existence, and, in general, the possible non-existence of something, could basically be represented in two ways: by means of a representation that "encodes" the explicit information that I do not exist, or by means of a representation that does not contain any information about me at all. In this latter case, the representation could be taken as the representation of the non-existence of something in case it can be somewhat "compared" with another representation that represents me (or something) as existing.[9]

[8] Properly speaking, this is not true: from the fact that a proposition is a representational entity does not follow that a set of propositions is a representational entity. World stories, however, are sets of propositions *with certain features*: a world story is able to represent actuality as being in a certain way, because it is a maximal and consistent collection of propositions about actuality. Maximality and consistency grant world stories their representational powers.

[9] Maybe, there are ways of representing implicitly the non-existence of an object without any need to compare such representation with another one that represents the object as existing. In such cases, one might say that the non-existence of something, while implicit,

For example, assume that there actually exist only three objects: John, Sam and myself and let us stipulate that figure 1 below is a representation of the actual situation, as far as the existence of objects is concerned.

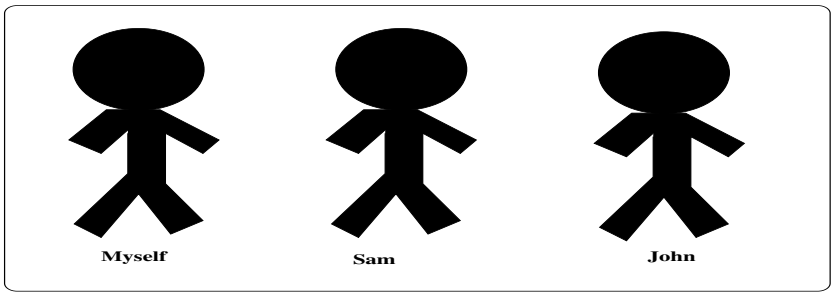



*Figure 1*

Now, assume further that being black in colour means, in the representation, that an object exists and being white in colour that it does not. Figure 2 below could then be taken as a representation that represents my non-existence *explicitly*:

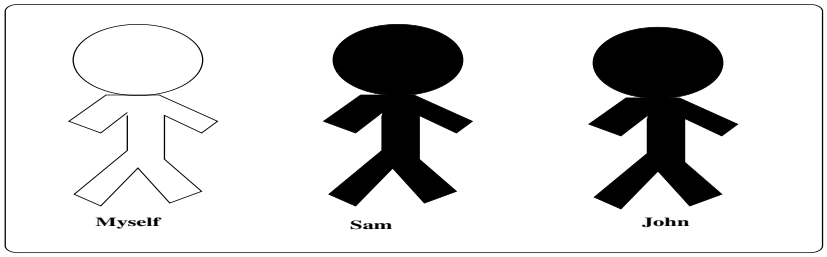



*Figure 2*

The following could instead be taken as a representation that represents my non-existence *implicitly*:

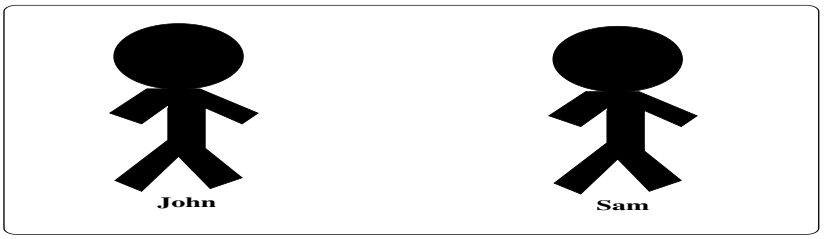



*Figure 3*

In the first case, I could gather the information that figure 2 is a representation of my non-existence *from within* the picture: my non-existence is represented by an intrinsic (graphical) property of the representation; knowing that white figures represent non-existing objects is sufficient to conclude that 2 is a representation that represents an object, namely myself, as non-existing. In the second case, I

it is also a "part" a "feature", or a "property" of the representation. In this article, I will only consider the "comparative" way of representing the possible non-existence of an object, because it is the one that corresponds to the notion "truth at" to be discussed below.

could only gather the information that figure 2 is a representation of my non-existence *from the outside*: the information is somewhat "inferred by comparison" (an expression to be qualified below) with the actual situation represented in figure 1.

The propositional counterparts of the three figures above are, respectively, the following three sets of singular propositions (I am representing singular propositions between corners and composed by objects and properties and I am assuming that existence is a genuine property of objects):

- $s^*$ = {<Sam, existence>, <John, existence>, <Myself, existence>}
- $s_1$ = {<Sam, existence>, <John, existence>, <Not: Myself, existence>}
- $s_2$ = {<Sam, existence>, <John, existence>}

Properly speaking, $s^*$, $s_1$ and $s_2$ are not world stories, because they are not maximal entities. We can assume, however, that a notion of "truth in a set of propositions $S$" could be defined for these non-maximal representational entities along the lines of the notion of "truth in a world story $S$" given in section 2.

According to this definition, we can conclude that the proposition that I do not exist is true in $s_1$, because the singular proposition <Not: Myself, existence> belongs to $s_1$.

What about the same proposition in $s_2$? We know from the recursive clauses (and the definitions of maximality) that an atomic proposition or its negation has to belong to a set of propositions $s$ to be true in $s$. In the case of $s_2$, neither the proposition that I exist nor the proposition that I do not exist belong to $s_2$, thus neither of them should be taken as true in $s_2$. So none of the definitions above could be of any help here.

The world stories theorist, however, believes that $s_2$ could be taken as an implicit representation of my non-existence and that the way in which the proposition that I do not exist could somewhat be "inferred by comparison" from $s_2$ can be captured in a semantically robust way. In order to do this, the preliminary thing is to assume that $s_2$ is related to $s^*$ in a relevant way, by representing a possibility for (what is represented in) $s^*$. She then introduces a novel relativized notion of truth: *truth at*, whose main feature is that of not requiring a proposition to belong to or to be true in a set of propositions $s$ or a world story $S$ to be true *at* $s$ or at $S$.

The intuitive motivation behind the distinction between, "truth in" and "truth at" a world story is well explained by Fine when he presents the distinction between two notions of propositional truth, the *inner* and the *outer*:

> One should distinguish between two notions of truth for propositions, the inner and the outer. According to the outer notion, a proposition is true in a possible world [in our case in a world story] regardless of whether it exists in that world; according to the inner notion, a proposition is true in a possible world only if it exists in that world. We may put the distinction in terms of perspective. According to the outer notion, we can stand outside a world and compare the proposition with what goes on in the world in order to ascertain whether it is true. But according to the inner notion, we must first enter the proposition into the world before ascertain its truth (Fine 1985: 163).

Inner truth corresponds to the notion of truth in, outer truth to the notion of truth at. The proposition that I do not exist is true in $s_1$, because it belongs to $s_1$

(it exists in $s_1$), the proposition that I do not exist is not true in $s_2$ (because it does not belong there), but it is nonetheless true at $s_2$, because its truth might be determined by comparing $s_2$ with $s^*$, the true world story.

The notion of "truth at" is thus very important for the actualist: it allows one to say that a given proposition is true with respect to a world (the proposition that I do not exist is true with respect to $s_2$) without assuming the existence of such proposition and of the objects the proposition is about in such a world.

According to Adams (1981: 23), the two basic principles regulating the notion of "truth at" are the following:

**Truth-at 1:** Every proposition Γ that is true in a world story $S$ is true at $S$.

**Truth-at 2:** In case at least one of $a_1, \ldots, a_n \notin S$, $\neg\Phi^n a_1, \ldots, a_n$ is true at a world story $S$ (where $\Phi^n$ is a primitive predicate).

From **Truth-at 1**, it follows that if an actual object $a$ belongs to a world story $S$, then everything that is true *in* $S$ of $a$ will be also true at $S$ of $a$; from **Truth-at 2**, it follows that, if $a$ does not belong to $S$, then the negation of every singular proposition about it will be true at $S$. In particular, if $a$ does not belong to a world story $S$, the proposition that $a$ does not exist will be true at $S$. On the basis of **Truth-at 2**, we may claim that the proposition that I do not exist is true at $s_2$, because I am not a constituent of $s_2$. By means of the notion of "truth at", we may finally make sense of the idea that $s_2$ represents my non-existence implicitly, that from $s_2$ one could infer my non-existence.

The situation is thus the following. There are two ways of representing the possible non-existence of an actual object within the world stories approach. One is the explicit way for which the dear old notion of "truth in" is enough, the other is the implicit way for which the novel notion of "truth at" needs to be introduced. My claim will be that both ways are problematic for the world stories approach. In particular, I will show that the notion of "truth at" is incompatible with the notion of maximality used within the world story approach (be it maximality* or simple maximality).

Before proceeding, I wish to propose a more rigorous presentation of the notion of "truth at", by giving a complete recursive definition of it, something that it is rarely found in the literature.[10]

As we have seen, the basic idea in the notion of "truth at" is that a proposition Γ can be true at a world story $S$ without existing in $S$. This is especially plausible within an actualistic framework where possible worlds and propositions are all actual entities. A recursive definition of "truth at" should be done in such a way that none of its clauses entail the existence of the corresponding proposition at the relevant world story (note, by comparison, how instead the first clause of the recursive definition of "truth in" immediately entails the existence in $S$ of the atomic propositions that are true in $S$). We will then say that an atomic proposition $\Phi a$ is true at a world story $S$ iff $S$ *represents** $a$ to be $\Phi$, where the representing* of $\Phi a$ by $S$ does not imply the existence of $\Phi a$ in $S$. Representing* is, of course, a new primitive, but it could stand for the explicit or the implicit way of representing mentioned above. In case a world story $S$ represents implicitly $\Phi a$ to be true, this

[10] Cf. for example, Turner 2005 where four necessary conditions for the notion of "truth at" are individuated, but a recursive definition is not given. King 2007: 83 is an exception and what follows is partially inspired by his view.

means that the extension of Φ could somehow, be read off and it could be determined that $a$ falls in the extension of Φ with respect to $S$. We do not need to be specific about the ways in which we determine the extension of a predicate Φ with respect to a world story $S$ (it might depend on specific features of the predicate), but this determination might occur by comparison with another world story (typically, the actual world story). Note that the fact that, in $S$, $a$ is represented* as falling in the extension of Φ does not entail, *per se*, that the proposition $\Phi a$ exists in $S$ (nor that $a$ exists in $S$). The recursive definition of "truth at" could thus be something like this:

- any atomic proposition $\Phi^n(a_1, \ldots, a_n)$ is true at $S$ if and only if $S$ represents* $a_1, \ldots, a_n$ as being in the extension of $\Phi^n$ ;
- $\neg\Gamma$ is true at $S$ if and only if Γ is not true at $S$;
- $(\Gamma \vee \Delta)$ is true at $S$ if and only if Γ is true at $S$ or Δ is true at $S$;
- $\forall x_1, \ldots, x_n\, \Phi^n(x_1, \ldots, x_n)$ is true in $S$ if and only if for any $a_1, \ldots, a_n$, $\Phi^n(a_1, \ldots, a_n)$ is true at $S$;
- □ Γ is true in $S$ if and only if for any world story $S_i$, Γ is true in $S_i$;
- for any other proposition Ψ, Ψ is not true in $S$.

Now let us see why the world stories approach, even armed with the brand new notion of "truth at a world story" cannot represent, in an actualistically acceptable way, the possible non-existence of an actual object.

### 4.1 Problems with the Explicit Way of Representing my Possible Non-Existence

The explicit way of representing my possible non-existence is problematic because *it is not actualistically acceptable*. *Prima facie*, this might sound quite surprising. Reductive kinds of actualism only require that the entities that go proxy for possible worlds be actual entities. From this point of view, $s_1$ seems to be perfectly acceptable: it only contains propositions whose objectual components are actual objects. Why then the explicit way of representing my possible non-existence should be problematic for an actualist?

To understand why, we need to introduce four ideas.

1. The first idea is that there are some connections between the modal status of a proposition (or a set of propositions) and certain counterfactual claims; in particular, if a proposition Γ is merely possible (*i.e.*, possible and not actually true), then the following principle seems to be true:

   (Poss-Count) If Γ is (merely) possible, then, had things gone differently, then Γ could have been true.

   On the other hand, if Γ is necessary, the following principle seems to be true:

   (Nec-Count) If Γ is necessary, then, no matter how things could have gone, Γ would have been true.

   These principles connect possibility and necessity claims to counterfactual claims and testify for their intimate logical relationships. In particular,

(Nec-Count) could be used, and indeed it has been used, to define metaphysical necessity in counterfactual terms.[11]

2. The second idea is that a false world story—a world story containing at least one false proposition—has the modal property of *possibly being the true world story*, where the true world story is the only world story containing all and only true propositions. Given (Poss-Count), the possible actuality of a world story $S$ implies that, had things gone differently, $S$ could have been the true world story. Call this thesis the *possible actuality of* (*false*) *world stories*.

3. The third idea is that, for the actualists, the claim that merely possible objects do not exist is *necessary*, not contingent. They not only believe that merely possible objects do not exist, but also that it is necessarily so, namely that there *could not have been* merely possible objects. Call this claim the *necessary non-existence of the non-actual*.

4. The fourth idea is that, for the actualists, the claim that there are no non-actual objects is usually taken to imply that that there are no *facts*, i.e., true propositions, about non-actual objects. Non-actual objects have no properties and are not involved in any relation with other actual or non-actual objects. Within the Russellian conception of proposition, the absence of facts about non-actuals is represented by the absence of propositions containing non-actuals as constituents. This thesis is usually called *serious actualism*.

The combination of the necessary non-existence of the non-actual (3) and serious actualism (4) implies that there could not be any facts about non-actual individuals and, given (Nec-Count), this claim should be understood as the claim that, in whatever way things could have gone, there would have been no facts about non-actual individuals.

Consider now $s_1$. Given that we have stipulated that $s^*$ is the true world story, $s_1$ is a false world story, because it contains the false proposition that I do not exist. Being a false world story, $s_1$ has the property of being possibly actual (by the possible actuality of false world stories) and then it is possible that $s_1$ is the true world story. Given (Poss-Count), from this it follows that had things gone differently, $s_1$ could have been true. There is then a counterfactual circumstance $C$, where $s_1$ is true. But what would have happened, had $C$ been the case? Well, it would have happened that I would not have existed, but—what is more important—that there would have been a fact about me, a non-existent object. Had $s_1$ been the actual world story (i.e., had $C$ been the case), a singular proposition having me as a constituent would have been true. Given 3, however, this cannot happen. In no counterfactual circumstance (and thus neither in $C$), there should be a true proposition having a merely possible object as a constituent and I would have been a merely possible individual, had the counterfactual circumstance $C$ be true. This situation is thus not actualistically acceptable. Given the necessary non-existence of non-actuals and (Nec-Count), no matter how things could have gone, there should be no facts about non-actual individuals. However, had things gone

[11] Cf. Williamson 2008: 159. To properly express the "no matter how things could have gone" in (Nec-Count) we probably need quantification into sentence position. The right-hand side of (Nec-Count) becomes $\forall p\ (p \mapsto \Gamma)$, where "$\mapsto$" is a would-counterfactual, which is provably equivalent to $\neg\Gamma \mapsto \Gamma$. This formula is used by Lewis 1973: 22 to define necessity.

the way $s_1$ represents them to go, there would have been a fact about a non-actual individual, namely myself.

The very same conclusion should be reached also by another route, namely by reflecting on some features of the theory of propositions behind the world stories approach. As I said, a world story theorist is (typically) an objectualist and a structuralist about propositions. Not simply so, however; indeed, she is also an *essentialist* about his structuralism and objectualism, for she believes that the identity of propositions is rooted in their constituents and components. One of the advantages of the Russellian conceptions of propositions over its non-structural (propositions as unstructured sets of circumstances of evaluations) and non-objectual (propositions as composed by intensional entities) competitors is that it allows for a finer-grained individuation of them; in particular, necessarily equivalent propositions may be distinct; the necessarily equivalent propositions expressed by $\forall x\ (x = x)$ and $\forall x\ \exists y\ (x = y)$ may be distinguished for structural reasons, the necessarily equivalent propositions expressed by "Socrates = Socrates" and "Plato = Plato" may be distinguished for objectual reasons. Essential objectualism, essential structuralism and actualism imply a thesis usually called *existentialism*:

**Existentialism:** if a proposition $\Gamma$ exists and it is about an object $a$ (it has $a$ as a direct constituent), then, had $a$ not existed, neither $\Gamma$ would have existed.

The thesis of existentialism implies that in a situation where I do not exist, no proposition having me as a constituent exists, neither the proposition that I do not exist. Being an existentialist, then, the world stories theorist cannot accept that my possible non-existence be explicitly represented by a set of propositions like $s_1$. No world story can have as a member a proposition to the effect that some object $a$ does not exist and therefore, the explicit representation of my possible non-existence is not a viable option for the world stories theorist.

Note that the first way of showing that the explicit way of representing my possible non-existence is not actualistically acceptable—the one based on (1)-(4)—has a clear advantage over the second way, based on Existentialism. The latter is explicitly grounded on principles belonging to a particular theory of propositions, while the former is based on general modal principles. Even though some of these principles (e.g., serious actualism) may be better understood within a Russellian setting, their plausibility is independent from such a propositional setting.

### 4.2 Problems with the Implicit Way of Representing my possible Non-Existence

The implicit way of representing my possible non-existence is problematic because it clashes with the definition of maximality in use (be it maximality* or maximality). Under maximality*, a world story $S$ is maximal* if and only if *for any actual object* $a$ and for any atomic predicate (or property) $\Phi$, either $\Phi a$ belongs to $S$ or its negation belongs to $S$. Under maximality, a maximal world story will contain (among others) any atomic proposition (or its negation) about every actual object. The result, in both cases, is that every world story will contain lots of propositions directly about *every* actual object; every world story will contain as many singular propositions about $a$ as many atomic predicates. In particular, for

any world story $S$ and actual object $a$, $S$ will contain either the proposition that $a$ exists or the proposition that $a$ does not exist.

This situation does not even allow for an implicit characterization of my possible non-existence. For this reason, the notion of "truth at"—explicitly designed for such a purpose—becomes completely useless. As we have seen in section 2, semantic notions such as "truth in" and "truth at" enter the scene only *after* maximal and consistent sets have been generated. We first define the notion of "world story" as a maximal consistent set of actually existing propositions and only then we can recursively define the notion of "truth in a world story $S$" or "truth at a world story $S$". Both notions simply presuppose the existence of world stories.

But the way in which world stories are built deprives the notion of "truth at" of its very rationale. Given a world story $S$ that represents my non-existence, due to the maximality of $S$, $S$ will contain the proposition that I do not exist and thus it is already true *in* $S$ that I do not exist.

The notion of "truth at", however, was supposed to help us just in representing possible situations where I do not exist without committing us to the existence of the proposition that I do not exist. But with the definitions of maximality in use (maximality and maximality*), the information that I do not exist according to a certain world story $S$ is something that could be gathered already by means of "truth in": the proposition that I do not exist is a member of a world story $S$ representing my possible non-existence $S$ and we know, from the recursive definition of "truth in", that every proposition that is a member of $S$ is also true in $S$. Introducing a further notion of "truth at" would be, at this point, simply superfluous.

From a general point of view, the problem is that what the official definitions of maximality produce is a complete description of a possible development of the actual world *from our perspective*; what would be needed is instead a complete description of a possible development of actuality *from the perspective of those* (*actuals*) *that would have existed according to such a possible development*.

The world stories theorist thus faces a dilemma: either she gives up the requirement that her world stories be maximal* or maximal sets of actually existing propositions or she gives up the notion of truth at a world story. In this latter case, however, false world stories will not be actualistically acceptable entities, because they would contain propositions about objects that would exist had one of those world stories become the true world story. In the former case, world stories could not be taken as total descriptions of how things could have gone and therefore they could not be taken as the right kind of entities to reduce possible worlds, not even those where only actual objects exist.

## 5. Local Maximality

My aim in this section is to help the world stories theorist to escape the dilemma presented above by proposing an alternative, and more plausible, conception of maximality, a notion of maximality that could be used in conjunction with the notion of "truth at".

In order to do this let me firstly emphasize again the distinction between two ways in which the notion of "complete description of an alternative course of actuality" could be understood. When I claim that some representational entity is a complete description of an alternative course of actuality, what I claim seems to be ambiguous between two readings:

- a complete description of actuality with respect to an alternative course of it;
- a complete description of an alternative course of actuality.

The two readings are obviously connected with the two ways, mentioned above, in which one could represent my possible non-existence. The first reading corresponds to the problematic notion of maximality in use so far. According to such a conception, a complete description of an alternative course of actuality is to be done *from the point of view of actuality*. If the actual objects were Sam, John and myself, a complete description of an alternative course of actuality in this sense would be a description that always tells us explicitly everything about Sam, John and myself. For any atomic predicate $P$, this kind of description tells us explicitly whether Sam, John and myself satisfy the predicate $P$ or not with respect to the possible situation to be described. In case the description describes an alternative course of actuality in which I do not exist, the description will tell us explicitly that I do not exist according to this alternative course and it will contain a proposition to the effect that I do not exist. Call this kind of maximality *global maximality*. Both maximality* and maximality are kinds of global maximality.

The second reading corresponds to a conception according to which the descriptions of alternative courses of events should be complete *from the point of view of these courses of events*. In such a way, descriptions are generated that are complete only with respect to the actual objects that would have existed, had the descriptions being true. Call this notion of maximality *local maximality*.[12]

I propose to characterize local maximality by means of the notion of *actual object that would have existed had a certain (set of) proposition(s) been true*. The idea is to define maximality for a world story only with respect to those objects that would have existed, had that set of propositions been true.

In the case of a set of atomic propositions Γ, it is quite easy to determine what objects would have existed had Γ been true. Let us consider the simplest case, namely that of a set of atomic propositions $s_1$, whose only component is the atomic and singular proposition $<P, a>$, expressed by $Pa$. Assume that $<P, a>$ is actually false and ask yourself: what actual objects would have existed had $s_1$ been true? In this case the answer is quite easy and unequivocal: assuming that the property $P$ is a qualitative property (and so does not involve the existence of any individual), $a$ is the actual object that would have existed, had $s_1$ been true.[13]

[12] The basic idea behind the alternative notion of maximality that I am going to present has been already envisioned by Adams (1981: 23). He in fact recognized the need to give some limitations to the maximality of world stories: "*Intuitively, a world story should be complete with respect to singular propositions about those actual individuals that would still be actual if all the propositions in the story were true, and should contain no singular propositions at all about those actual individuals that would not exist in that case. For the propositions would not exist and therefore could not be true, if the individuals did not exist*". The problem, however, is that this limitation on maximality and the consequences of this limitation have never been explicitly worked out, either by Adams and, as far as I know, by any other world stories theorist.

[13] Admittedly, the notion of a qualitative property is far from being clear. A qualitative property is usually defined as a property whose linguistic formulation does not contain any referential device to individuals. If $\lambda x\Phi$ is the linguistic formulation of a qualitative property, then in order for the property to count as qualitative, $\Phi$ should not contain any individual constant or free variables (except from $x$). This definition is problematic in a number of ways: for a start, it might be extensionally wrong in that there might be inexpressible

If we want to represent a possible situation in which only $a, b$ and $c$ exist (whereas, in the actuality, there exist $a, b, c$ and $d$), our world story will be a set of consistent and maximal atomic propositions about $a, b$ and $c$, i.e., having only $a, b$ and $c$ as objectual components.

Things are not so simple in the case maximality is not defined only with respect to the atomic singular propositions. In such a case, the truth of a proposition Γ might be compatible with the existence of distinct sets of actual objects; hence, the answer to the question "What actual objects could have existed, had Γ been true?" turns out to be more difficult to answer.

Consider, for example, the proposition expressed by $\exists x\ (x \neq a)$ and assume that such a proposition is a member of a consistent set of propositions $s_{ii}$ whose other member is the proposition expressed by $Pa$, namely <$P$, $a$>. Assume that the set of actual objects is $\{a, b, c\}$. Now, what actual objects would have existed had $s_{ii}$ been true?

The truth of the proposition expressed by $Pa$ requires the existence of $a$ (assuming, again, that $P$ is qualitative), but the truth of the proposition expressed by $\exists x(x \neq a)$ is compatible with the existence of distinct sets of actual objects, namely $@_{s_1} = \{a, b\}$, $@_{s_2} = \{a, c\}$, $@_{s_3} = \{a, b, c\}$. Had $s_{ii}$ been true, the actual objects that would have existed had $s_{ii}$ been true would have been either $@_{s_1}$ or $@_{s_2}$ or $@_{s_3}$. Call the distinct sets of actual objects compatible with the truth of a certain set of propositions, the $@_s$-sets.

Now, the idea is that we can answer unequivocally to the question "what actual objects are compatible with the truth of a certain set of propositions $s$?" only relatively to an $@_s$-set. The question "*what actual objects would have existed had s been true?*" should then be reinterpreted as the question "*what actual objects, relatively, to an $@_s$-set, would have existed, had s been true?*" With the expression "the actual objects that, relatively to $@_{s_i}$, would have existed, had $s$ realized" I will simply denote all the (actual) objects belonging to $@_{s_i}$.

With the notion of $@_s$-set at our disposal, we are now ready to define a notion of local maximality—that I will call $@_s$-maximality—for a set of actually existing propositions $s$.

I will use $\Gamma_i$ or $\Delta_i$ to refer to the propositions Γ and Δ such that all their objectual components (if there are any) are elements of the set $i$. I will use an expression like $P^n$ to refer to any primitive property and I will use an expression like $\Phi_i^n$ to refer either to a primitive property or to any non-primitive property Φ such that its only objectual components (if there are any) are elements of the set $i$.

qualitative properties. Furthermore, a purely "syntactic" criterion does not account for predicates that could contain "indirect" semantic relations to individuals or places (e.g., "Hellenic"). One could say that a property is qualitative in case the predicate that expresses it does not semantically "involve" individuals, but, admittedly, even this formulation is not very precise. As far as this paper is concerned, I will rest content with this generic formulation. Cf. Williamson 2013: 271. Note that, for my purposes, I do not need to restrict myself on or to have any special commitment to qualitative properties as far as it is clear what objects the truth of a proposition implies. In case $Pa$ is a false proposition and $P$ non-qualitative, the objects that would have existed had $Pa$ been true, would have been $a$ and all the objects whose existence would have been implied by $P$. The only constraint on properties is that they do not involve any semantic relations to merely possible objects, but this is already part of the actualist spirit of the world stories approach.

A set of propositions $s$ is $@_{s_i}$-maximal, for some $i$, if and only if

- for any $n$-ary property $P^n$ and for any actual object $a_1, \ldots, a_n \in @_{s_i}$, either $P^n(a_1, \ldots, a_n) \in s$ or $\neg P^n(a_1, \ldots, a_n) \in s$;
- $\neg\Gamma_{@_{s_i}} \in s$ if and only if $\Gamma_{@_{s_i}} \notin s$;
- $(\Gamma_{@_{s_i}} \vee \Delta_{@_{s_i}}) \in s$ if and only if $\Gamma_{@_{s_i}} \in s$ or $\Delta_{@_{s_i}} \in s$;
- $\forall x \Phi_{@_{s_i}}(x) \in s$ if and only if for any $a \in @_{s_i}$, $\Phi_{@_{s_i}} a \in s$;
- for any other proposition $\Psi$, $\Psi \notin s$.

With the notion of $@_{s_i}$-maximality at hand, we can now define a world story as follows:

If $s$ is a set of propositions, $s$ is a *world story* if and only if $s$ is consistent and, for some $i$, $@_{s_i}$-maximal.

The new conception of maximality allows us to generate world stories where my possible non-existence is represented simply by the lack of any proposition having me as a constituent; such world stories are also locally maximal in the sense of being complete descriptions of alternative courses of actuality. A world story representing an alternative course of actuality where I do not exist is a set of propositions that do not have me as an objectual constituent. The notion of local maximality now ensures that I do not belong to such a world story, because I am not belonging to any of the @-sets representing a possible development of actuality in which I do not exist. @-sets select only objects that would have existed, had certain propositions been true and I would not have been selected, had the proposition that I do not exist been true.

The world story theorist can now profitably use the notion of "truth at" in order to represent implicitly the possible non-existence of an actual object. My possible non-existence is true at a world story that does not have me as a constituent. The notion of "truth at" can now do the work it was designed to do, namely that of allowing us to make comparisons between two locally maximal and consistent world stories. By means of the notion of local-maximality, the framework of world stories is now compatible with the introduction of the notion of "truth at" and can represent, in an actualistically acceptable way, the possible non-existence of actual objects.

## 6. Conclusion

My conclusion is that the world stories theorist can represent the possible non-existence of actual objects if she abandons the notion of global maximality and uses instead the notion of local maximality (@-maximality). In this paper, I have shown why global maximality generates world stories that the actualist should not accept and, furthermore, why it makes the notion of "truth at a world story" useless. But a working notion of "truth at" is essential for the world stories approach, because it is only by means of such a notion that the world stories theorist will be able to represent the possible non-existence of an actual object in an actualistically acceptable way, namely without assuming the existence of world stories containing singular propositions about objects that would not exist, had that world story been the true world story. The notion of local maximality is thus needed to represent, by means of the notion of "truth at", my possible existence

in an actualistically acceptable way and solve the problem of the possible non-existence of actuals for the world stories approach.

## References

Adams, R.M. 1974, "Theories of Actuality", *Nous*, 8, 211-31; repr. in Loux 1979, 190-209.

Adams, R. M. 1981, "Actualism and Thisness", *Synthese*, 49, 3-41.

Carnap, R. 1947, *Meaning and Necessity*, Chicago: Chicago University Press.

Divers, J. 2002, *Possible Worlds*, London: Routledge.

Fine, K. 1977, "Prior on the Construction of Possible Worlds and Instants", in *World, Times and Selve*s, London: Duckworth, 116-61; repr. in Fine 2005, 133-74.

Fine, K. 1980, "First-Order Modal Theories", *Studia Logica*, 39 (2-3), 159-202.

Fine, K. 1985, "Plantinga on the Reduction of Possibilist Discourse", in Tomberlin, J.E. & Van Inwagen, P. (eds.), *Alvin Plantinga*, Dordrecht: Reidel, 145-86.

Fine, K. 2005, *Modality and Tense*, Oxford: Oxford University Press.

King, J. 2007, *The Nature and Structure of Content*, Oxford: Oxford University Press.

Kripke, S.A. 1980, *Naming and Necessity*, Harvard: Harvard University Press.

Lewis, D.K. 1973, *Counterfactuals*, Oxford: Blackwell.

Linsky, B. & Zalta, E. 1994, "In Defence of the Simplest Quantified Modal Logic", *Philosophical Perspectives*, 8, 431-58.

Loux, M.J. (ed.) 1979, *The Possible and the Actual*, Itacha: Cornell University Press.

Menzel, C. 2016, "Actualism", in Zalta, E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2016 ed.). http://plato.stanford.edu/archives/21sum2016/entries/actualism/.

Turner, J. 2005, "Strong and Weak Possibility", *Philosophical Studies*, 125 (2), 191-217.

Williamson, T. 2008, *The Philosophy of Philosophy*, Oxford: Blackwell.

Williamson, T. 2013, *Modal Logic as Metaphysics*, Oxford: Oxford University Press.

# Natural Properties Do Not Support Essentialism in Counterpart Theory: A Reflection on Buras's Proposal

*Cristina Nencha*

*University of Turin*

*Abstract*

David Lewis may be regarded as an antiessentialist. The reason is that he is said to believe that individuals do not have essential properties independent of the ways they are represented. According to him, indeed, the properties that are determined to be essential to individuals are a matter of which similarity relations among individuals are salient, and salience, in turn, is a contextual matter also determined to some extent by the ways individuals are represented.

Todd Buras argues that the acknowledgment of natural properties in counterpart theoretic ontology affects Lewis's theory with regard to essentialism. Buras's reasoning is appealing. He claims that, since natural properties determine the existence of similarity relations among individuals that are salient independent of context, Lewis can no longer be claimed to be an antiessentialist.

The aim of this paper is to argue, against Buras, that if counterpart theory was antiessentialist before natural properties were taken into account, then it remains so afterwards.

*Keywords*: David Lewis, Counterpart Theory, Essentialism, Natural Properties, Todd Buras.

## 1. Introduction

David Lewis may be regarded as an antiessentialist, for it is said that, according to him, individuals do not have real essential properties, that is they do not have essential properties independent of the ways they are represented—namely, conceived or described. This is the case because: (a) the properties that are determined as essential to individuals are a matter of which relevant counterparts they have, (b) the relevant counterparts that individuals have are a matter of which similarity relations are salient, and (c) salience is a contextual matter also determined to some extent by the way individuals are represented.

*Argumenta* 2, 2 (2017): 281-292
ISSN 2465-2334


DOI 10.23811/46.arg2017.nen

In his *New Work for a Theory of Universals*, Lewis defends the view that there are both abundant and sparse properties.[1] Among the sparse properties there is a group of natural properties that marks out the genuine qualitative similarities and differences between individuals.[2]

Todd Buras, in his *Counterpart Theory, Natural Properties and Essentialism*, argues that, if Lewis accepts both counterpart theory and natural properties, he can no longer be classified as an antiessentialist. This is because the natural properties determine the existence of similarity relations among individuals that are salient, and therefore relevant, independent of the ways those individuals are conceived or described. If such similarity relations are obtained, Buras claims, individuals have real essential properties.

The aim of this paper is to argue, against Buras, that the implications of counterpart theory for essentialism are not altered by the acknowledgement of natural properties. If counterpart theory was antiessentialist without natural properties, then it remains so after natural properties are taken into account.

## 2. David Lewis's Antiessentialism

There are different characterizations of essentialism.

One might say that essentialism is the doctrine holding that at least some individuals have both essential and accidental attributes. According to this characterization, for instance, anyone who believes that, for every individual *a*, all of *a*'s attributes are essential to it, is defined as an antiessentialist.

Alternatively, one might characterize essentialism as the thesis that at least some individuals have some essential properties, so that someone who was an antiessentialist by the earlier criterion would now count as an essentialist.

One might also want to distinguish between trivial and nontrivial essential properties. Trivial essential properties are properties such as being either *P* or non-*P*, for any property *P*.[3] Then, she might take essentialism to be the doctrine that at least some individuals have some nontrivial essential attributes. Thus, anyone who believes that individuals have only trivial essential attributes is regarded as an antiessentialist.

I neither need to say that this exposition is exhaustive nor to choose which kind of characterization is the best definition of essentialism. However, for the sake of argument, let us take essentialism as the latter thesis, which argues that some individuals have at least some nontrivial essential attributes. So far, then, commitment to essentialism is simply a matter of being prepared to say, without

[1]Abundant properties are highly disjunctive properties, therefore, they are undiscriminating. By contrast, sparse properties are highly specific and characterize things completely and without redundancy. See Lewis 1983: 346-47; 1986: 59-61.

[2] Throughout this paper I will use "lazy" talk about properties. From Lewis's perspective, indeed, properties cannot explain anything. He is a class nominalist; he identifies properties with classes of particulars, and belonging to one class is a primitive fact that cannot be explained further.

[3] In the example, the triviality of the property of being *P* or non-*P* relies on the fact that this property belongs to all things. For attempts to establish which other properties count as trivial essential, see Marcus 1967 and Della Rocca 1996.

further explanation or characterization, that an individual has nontrivial essential properties. Nothing has been said about what is required for a property to be an essential property. Let us call this conception of essentialism "realistic-neutral essentialism".

There is a further requirement for a stronger, metaphysically more robust conception of essentialism.[4] Given an individual *a* and an attribute *P*, "*a*'s being essentially *P*" is a matter independent of the ways in which we conceive or describe *a*. Independently of the ways an individual is represented, there is a fact of the matter about its being essentially something. Let us call this stronger conception of essentialism "realist essentialism". According to realist essentialism, individuals have thus—nontrivial—real essential properties.

There is thus another way to be antiessentialist, that is to deny that individuals have real essential properties. It is in exactly this sense that David Lewis ought to be counted as an antiessentialist. Lewis is thus characterized as an antiessentialist, precisely when people have in mind realist essentialism.

Let us look at the reason for this.

As far as essentialism is concerned, we are interested in *de re* modality.[5] From Lewis's perspective, *de re* modality is explained through counterparts (Lewis 1968; 1986).

Let us consider sentence type 1:[6]

1. *a* is essentially human.

According to counterpart theory, 1 is true if and only if—hereafter, iff—every relevant counterpart of *a* is human.

The general form of the truth-conditions for an essentialist sentence type is thus incomplete: it needs to be completed with the input of a relevant counterpart relation.[7]

A counterpart relation between two individuals is any relation of similarity between them; counterparts of *a* are simply any things that are similar in any respect and to any degree to *a*. There is then the further question of which counterparts of *a* are relevant; *b* is a relevant counterpart of *a* iff *b* is similar enough to *a* under relevant respects.

It is a matter of context which respects of similarity are salient and which grades of similarity are enough under such respects. The relevant counterparts of *a* are therefore determined to a large extent by the contexts in which 1 is produced and evaluated.

[4] This further condition is generally attributed to Quine. See Quine 1953a; 1953b.

[5] For Lewis, questions of essentialism are at one with questions of necessity *de re*. This is in common with many philosophers—like Quine, Kripke, and Marcus—but not with most philosophers after Kit Fine who would distinguish the two. See Fine 1994.

[6] I use the distinction type-token in order to underline the fact that, according to counterpart theory, and as will be shown, the logical form of an essentialist sentence is incomplete. This completion happens only at the level of specific tokens of that sentence.

[7] For instance, a token of 1 might be true iff *a*, *b* and *c* are human, while a different token of the same sentence type might be true iff *a* and *d* are human. This is because different tokens of the same essentialist sentence type can evoke different relevant counterparts: in the former case *a*, *b* and *c* are the relevant counterparts of *a*, while in the second case *a* and *d* are *a*'s relevant counterparts.

Counterpart theory thus gives complete truth-conditions only for specific tokens of 1. In other words, in order to have truth-values for essentialist claims about *a*, we need to know which of *a*'s counterparts are relevant, and this is determined for the greater part by the contexts in which the essentialist claims are uttered.

According to Lewis, the interests and intentions of a speaker and an audience, background information, the standards of precision, the presuppositions, spatiotemporal location of utterances, norms of charitable interpretation, and objective salience are among the contextual factors that help to select the relevant counterparts of individuals (Lewis 1979; 1980). What helps to select the counterparts of individuals that are relevant in a particular context, among other factors, are thus also the ways that those individuals are conceived or described.

Different tokens of the same essentialist sentence type about *a* might be produced and evaluated in different contexts and they can thus evoke different relevant counterparts according to those different contexts. Some tokens of the given sentence type might thus be true, and some others might be false. The properties that are determined as essential to an individual might therefore change from one context to another and, sometimes, all that changes from one context to another is our way of representing that individual.

For instance, there are at least two respects of similarity that we might take into account as relevant when we evaluate 1. We might take personhood as a relevant respect of similarity; alternatively, we might count bodyhood as a relevant respect of similarity. Which of these is relevant is a contextual matter. If the context in which the claim is produced assigns great weight to the former respect of similarity, then we see that *a*'s relevant counterparts are persons, since only persons can resemble another person with respect to personhood enough to be the relevant counterparts. Otherwise, if the context places great stress on the latter respect of similarity, we find as *a*'s relevant counterparts its bodily counterparts (Lewis 1971: 208). In the first context of utterance, the truth-conditions of 1 are completed by the input of personhood counterparts, while in the second context of utterance they are completed with regard to a different token of 1 by the input of bodyhood counterparts. According to the latter context, but not according to the former, it therefore might be false that *a* is essentially human. What changes from the first context to the second might be the way *a* is conceived or described. For instance, the first context might be one in which we conceive or describe *a* as "that person", while the second context might be one in which we represent *a* as "that thing". The token of 1 uttered in the first context is therefore true, while the token uttered in the second context might be false, and this is so because the two different tokens evoke different relevant counterparts according to different contextual ways of representing *a*. Thus, it turns out that different properties might be determined as essential to *a* according to different contexts and, specifically, according to different ways of representing *a*.

To sum up, according to Lewis essentialist sentences types do not have constant truth-values. This is because different tokens of the same sentence type about *a* might have different truth-values. Indeed, they might be produced and evaluated in different contexts and thus evoke different relevant counterparts of *a.* Some tokens of the given sentence type might thus be true, and some others might be false.

Individuals might therefore have different *de re* modal attributes according to different contexts. What may change from one context to another is, among other

features, the ways in which they are represented. Hence, the properties that are determined to be essential to individuals are sometimes influenced by our ways of conceiving or describing those individuals. Accordingly, in my classification, Lewis ought to be seen as rejecting realist essentialism—even though he accepts realistic-neutral essentialism.[8]

## 3. Buras's Proposal

In *New Work for a Theory of Universals,* Lewis defends the view that there are both abundant and sparse properties. Among the sparse properties there is a group of natural properties that marks out the genuine qualitative similarities and differences in things. According to Lewis, an adequate theory of properties has to recognize an objective difference between natural and unnatural properties; moreover, preferably, this difference has to admit degrees, so that the most natural properties are the perfectly natural properties (Lewis 1983: 346, 347). Natural properties characterize things completely and without redundancy; they carve out the joints of nature and it is the business of physics to discover these natural properties (Lewis 1983: 365, 366).[9]

From Buras's perspective, in admitting natural properties Lewis accepts that relations of overall similarity obtain among individuals, independent of the ways

[8] Following the suggestion of an anonymous referee, I would like to specify one point. There is some important literature about counterpart theory and mereological essentialism—that is, the controversial thesis that fusions have their parts essentially. One of the most discussed questions in this debate is whether or not counterpart theory, when conjoined with composition as identity, entails mereological essentialism. For opposing answers to this question see Merricks 1999 and Borghini 2005. However, since this paper is not concerned with composition as identity, I will leave this matter untouched. In general, I take mereological essentialism to be a special case of the general essentialist thesis, in which we focus our attention on a particular attribute that individuals may or may not have essentially, that is the attribute of being composed of their actual parts. As for the other kinds of attributes, I take Lewis's stand to be the same: under some counterpart relations individuals have essentially the parts they actually have, while under some other counterpart relations this is not the case. In Lewis's words: "I myself think that some counterpart relations validate Mereological Essentialism and other equally legitimate counterpart relations do not" (Lewis 2001: 608). At any rate, for the purposes of this paper it does not matter which kinds of attributes individuals have or do not have essentially. Rather, what counts is whether or not, in Lewis's view, the acknowledgment of natural properties can make "*a*'s being essentially *P*" a matter independent of the ways in which *a* is represented, quite regardless of which kind of attribute *P* is.

[9] It is well known that there are many unsolved questions about how naturalness should be understood in Lewis's metaphysics. They concern, among other things, the logical status of the notion of the natural in Lewis's metaphysics—for a survey, see Taylor 1993; how Lewis accounts for degrees of naturalness—see for instance Sider 1995 and Nolan 2005; which the bearers of natural properties are—for a non-standard reading of this matter see Borghini & Lando 2011—and so on. In this paper I am not taking issue with any of the problems that have been raised about the Lewisian characterization of naturalness, however. For the purposes of Buras's proposal, and thus for the aims of this paper, the details of naturalness do not matter, as long as some properties are classified as natural properties and there is a scale of degrees of naturalness.

those individuals are conceived or described. His first move in that direction is to define overall perfect natural similarity as follows:

2. *a* is overall perfectly naturally similar to $b =_{df} a$ shares at least one of *b*'s perfectly natural properties, and there is no individual *c*, distinct from *a* and *b*, such that *c* shares at least one of *b*'s perfectly natural properties, and *c* shares more perfectly natural properties with *b* than *a*.

He further claims that whether any two individuals are overall perfectly naturally similar or not is determined independently of the contexts and, mainly, independently of the ways in which those individuals are represented. Indeed, determining which properties are perfectly natural and which perfectly natural properties individuals have does not depend on contexts and, more specifically, it does not depend on our ways of conceiving or describing individuals.

Buras's crucial move then is to suggest that, if relations of similarity obtain among individuals independently of context and, thus, independently of our ways of representing those individuals, then individuals have real essential properties; that is, they have essential properties independent of the way they are conceived or described. This is because the overall perfect natural similarity determines the maximally natural counterparts—hereafter, MN counterparts—of the individuals: *b* is a MN counterpart of *a* iff *b*, in its own world, is overall perfectly naturally similar to *a*. The MN counterparts of *a* are therefore counterparts of *a* determined by virtue of similarity relations to *a* which hold independent of context and, specifically, independent of how *a* is represented.

If there are MN counterparts of individuals, then, according to Buras, individuals have real essential properties:

3. The real essential properties of *a* =*df* all and only the properties shared by all of *a*'s MN counterparts.

Indeed, the essentiality of the properties *a* shares with all of its MN counterparts is independent of the ways that *a* is represented.[10]

[10] The reading of Buras's proposal that I pursue in the rest of this paper is as follows: given that counterpart theory must accept the existence of similarity relations that hold independently of the ways individuals are represented, such a theory must also accept that individuals have real essential properties. The reason for such a conclusion is that the essential properties of individuals are determined by virtue of such similarity relations. Therefore—to anticipate an argument provided in Section 5—a part of my strategy to rebut Buras's argument is to argue that, since according to Lewis's definition, *a*'s essential properties are determined by the relevant counterparts of *a*, Buras's proposal would have success only in the case in which, in every context, *a*'s MN counterparts were the relevant ones—given that only MN counterparts are determined by virtue of that kind of similarity relations—which is not the case. A referee pointed out that there might be another way to read Buras's proposal. According to this reading, Buras's definition 3 should be read as a stipulation. Buras's crucial move in order to show that counterpart theory is committed to realistic essentialism is thus to stipulate, by definition 3, that real essential properties are determined by MN counterparts. If this reading is right, then there would be no need to show that MN counterparts are not the relevant ones in every context, since they are, by stipulation, the ones that determine the essential properties. Moreover, there would be a clear sense in which counterpart theory was committed to realistic essentialism. Indeed, as will

Buras therefore claims that counterpart theory with natural properties commits one to realist essentialism (Buras 2006: 32-37).[11]

## 4. Sharing Properties in Lewis's View

The first point that I would like to make in reaction to Buras's argument is that from Lewis's perspective, almost all similarity relations among individuals obtain independently of the ways those individuals are conceived or described. Indeed, similarity is defined in terms of properties sharing. The fact that two individuals have some properties in common, that they are similar in some way, does not depend, in general, on our ways of conceiving or describing them. To be sure, however, in some special cases the fact that two individuals share a property does depend on how they are represented. For instance, two individuals can be similar because they both have the property of being thought of by me or of being imagined by me and so on. However, for the most part, the sharing of properties is independent of how the individuals are represented. Whether or not *a* is similar to *b* is therefore usually independent of our ways of conceiving or describing *a* and *b*.

It should thus be said that, when they are shared, natural properties are not only supposed to give rise to similarity relations that hold independent of the ways individuals are represented, because this is true of almost any similarity relation. Rather, natural properties are supposed to mark out similarity relations between individuals that are metaphysically privileged, since they are similarities by virtue of shared fundamental properties. That is, the salience of such similarity relations is absolute, meaning that it is not contextually determined and, mainly, it is not determined by our ways of representing individuals. Such similarity relations then determine the MN counterparts, namely the counterparts that are metaphysically salient.

I am in agreement with Buras as far as this account of Lewis's metaphysical commitments is concerned. Counterpart theory, with natural properties in its ontology, has to accept the existence of counterpart relations and, *a fortiori*, of similarity relations among individuals that are metaphysically salient.

be shown, it is true that the Lewisian recognizes that there are MN counterparts, and thus clearly accepts that some individuals have some real essential properties, as Buras defines them. If the reader prefers the latter interpretation of Buras's proposal, she can read the rest of this paper as an attempt to show that counterpart theorists would not be obliged to accept Buras's stipulation about real essential properties, according to which *a*'s essential properties are determined by *a*'s MN counterparts. Indeed, according to Lewis, by definition, *a*'s essential properties are determined by the relevant counterparts and it is not at all obvious that the MN counterparts are always the relevant ones. Moreover, to appeal to MN counterparts in the definition of what makes for essential properties—as will be argued in Section 5—would lead to unacceptable consequences from the Lewisian point of view. My conclusion would then be that Buras is not successful in showing that counterpart theory is committed to realistic essentialism, even according to the latter interpretation.

[11] Ghislain Guigon has advocated a similar conclusion. See Guigon 2014. The difference between the two approaches lies mainly in the arguments they employ in order to defend this conclusion. For instance, as Guigon himself notes, while Buras believes that shared perfectly natural properties are privileged respects of similarity, Guigon argues that the similarity of perfectly natural properties counts as well. At any rate, I think that many of the arguments in this paper might also apply to Guigon's proposal—especially my metaphysical objection in Section 7.

Where I disagree with Buras is in my rejecting the conditional proposition that he asserts—that is: if there are such similarity relations and counterparts, then individuals have real essential properties.

In my semantical objection to Buras, I will argue that MN counterparthood is not metaphysically determined as the salient kind of counterparthood for every token of a given essentialist sentence type. The truth-conditions for essentialist sentence types are thus still inconstant and influenced, sometimes, by considerations of how we represent individuals.

In my metaphysical objection to Buras, I will claim that MN counterparthood is not metaphysically salient for determining which properties deserve to be characterized as the essential properties of individuals. The essentiality of those properties is still explained by facts about representation.

## 5. Semantical Objection to Buras

Recall that, according to Lewis, the truth-conditions of an essentialist sentence type about *a* are inconstant because different tokens of the same type might have different truth-values, since they might be produced and evaluated in different contexts and thus evoke different relevant counterparts of *a* according to those different contexts. Therefore, the counterparts of *a* that are semantically salient—that is, those relevant for the characterization of the truth-conditions of the sentence type—are those relevant in the contexts in which tokens of that type are uttered. Sometimes, what determines the contextual relevance of some kind of counterparthood of *a* is precisely the ways that *a* is represented in the context at stake. The truth-conditions of essentialist sentences types about *a* are thus always sensitive to context and, sometimes, they are sensitive to the modes of representation of *a*.

What I want to argue is that the situation remains unchanged after natural properties are taken into account. Indeed, we should not take Buras's metaphysical narrative above, according to which the acknowledgment of natural properties implies the existence of metaphysically privileged similarity relations, to have any automatic semantic implications. That is to say, it is not at all obvious that such similarity relations are, in virtue of their metaphysical privilege, semantically salient in every context so as to institute the salient kind of counterparthood for every token of a given essentialist sentence type. Even though there are metaphysically relevant counterparts, the truth-conditions of essentialist sentences types are not susceptible to being completed by the input of these counterparts in every context, because such counterparts are not, invariably, the semantically salient cases in every context. They are semantically relevant only according to some contexts, such as when the scientific perspective is contextually relevant, so that some token of some essentialist sentence type is made true by fundamental similarities. They might also be semantically relevant in the contexts in which a speaker's intentions and thoughts are not determinate enough to select a counterpart relation, so that we let the world decide for them. However, they are not semantically relevant according to every context.

We should consider how Lewis's account of semantics fits with his general theory of interpretation. That general theory emphasizes the charity of truthfulness. The interpretation of a speaker on an occasion is—*ceteris paribus*—the better for making the speaker a truth-teller. According to Lewis, there is a rule of accom-

modation holding that "what you say makes itself true, if at all possible, by creating a context that selects the relevant features so as to make it true" (Lewis 1986: 251). This is also true in *de re* modal contexts. For instance, when Kripkeans make claims of essentiality of origins they speak truly in the context of their own speaking. "They make themselves right: their preaching constitutes a context in which *de re* modality is governed by a way of representing (as I think, by a counterpart relation) that requires match of origins" (Lewis 1986: 252). In that context, according to Lewis's general theory of interpretation, we are thus bound to project backwards, as it were, the kind of counterparthood that must be selected in order to make their essentialist statements true.

If metaphysically salient counterparts were inevitably those selected, it would have massively uncharitable effects. Indeed, both Kripkean essentialist claims, and many other essentialist statements, do not seem to be made true by fundamental similarities. They would thus turn out to be false in the contexts of their own utterances, contrary to Lewis's expectations. So for Lewis, given how the semantics fits with his broader theory of interpretation, there is no chance of accepting the inevitable semantic salience of those counterparts that are metaphysically salient.

We should therefore not take the existence of metaphysically relevant similarity relations to imply that the truth-conditions for a given essentialist sentence type are truth-conditions completed in every context by the input of counterparts whose relevance is independent of our ways of representing individuals. Even with the acknowledgment of natural properties, the truth-conditions for essentialist sentences types are still sensitive, sometimes, to considerations of how we conceive or describe individuals.

## 6. Buras's Reply to the Semantical Objection

Buras anticipated a similar objection.

He admits that the context still determines whether *de re* modal claims are to be evaluated against natural or unnatural facts. He also admits that *de re* modal claims are about natural facts only in some contexts.

His point, however, is that we should not confuse metaphysics with semantics. From his perspective, an antiessentialist is not someone who believes in the inconstancy of *de re* modal statements, rather it is someone who believes that there are no *de re* modal facts, or facts of the matter about the essential properties of individuals.

Buras's argument is supposed to show that counterpart theoretic ontology with natural properties is able to give rise to *de re* modal facts, even though *de re* modal claims are inconstant. In his opinion, he only needs to show that "one counterpart relation stands out from the crowd for the purposes of characterizing the modal properties of an object" (Buras 2006: 40), which will be sufficient and hold independently of the semantic matter regarding whether or not the MN counterparts are also semantically privileged in every context.

Buras thus takes the acknowledgment of natural properties to imply that, metaphysically speaking, the MN counterparts are privileged in determining which properties deserve to be characterized as essential properties of individuals. This holds, in Buras's opinion, even though the MN counterparts are not semantically privileged in every context; even though they do not determine the truth-

values of every token of a given sentence type. According to Buras, this is what makes a theory a realistic essentialist theory.

Now, granted that Buras's account of Lewis's metaphysical commitments does not have any automatic semantic implications—that is, granted that the MN counterparts are not semantically relevant in every context—I shall go on to argue that it does not even have the further metaphysical implications that Buras takes it to have: the MN counterparts do not make for the existence of facts of the matter about individuals' essential properties.

## 7. Metaphysical Objection to Buras

Two claims constitute my metaphysical objection. Firstly, by virtue of the Lewisian principle of recombination, there cannot be any facts of the matter about individuals' essential properties. Secondly, this thesis is not altered by the acknowledgment of natural properties and, hence, of metaphysically salient counterparts of individuals.

From Lewis's perspective, metaphysics can only establish which properties an individual has and which of these are natural. Which of these properties deserves to be characterized as being held as a matter of *de re* necessity is then not a question that can be approached within metaphysics alone. Which kinds of factors select the property of an individual as one that it has essentially, for Lewis, is a question about which relevant counterparts are selected. Relevance is a contextual matter that, sometimes, depends on our ways of representing the individual. In other words, metaphysics cannot establish that a property of an individual is one that it has essentially, because this is intrinsically a contextual matter that depends, sometimes, on our ways of representing the individuals.

Contrary to what Buras claims, even though there are metaphysically relevant similarity relations, there is thus no room for facts of the matter about an individual's essential properties, because it is the business of the context to establish which properties are determined to be essential to individuals. This cannot be the business of metaphysics; there is no property of *a*—not even one shared by all the MN counterparts of *a*—that is metaphysically selected as one that *a* has as a matter of necessity *de re*.

With the above in mind, metaphysically speaking, there are no counterparts of *a*—not even the MN counterparts—that are privileged for determining which of *a*'s properties deserve to be characterized as essential properties, precisely because there are no properties metaphysically privileged to be characterized as essential properties.

With the acknowledgment of natural properties, all that metaphysics can do is to put counterpart relations on different levels: only some play a role in characterizing metaphysically relevant facts of resemblance. Afterwards, by definition, it is a task of the context to select the relevant counterparts that determine which properties an individual has as a matter of necessity *de re*.

The implications of counterpart theory for essentialism are not altered by the acknowledgement of natural properties; the existence of metaphysically privileged similarity relations does not change the fact that metaphysics, in principle, cannot establish which properties are determined to be essential to individuals. There are thus no counterparts that stand out from the crowd, metaphysically speaking, for determining which properties deserve to be characterized as essential properties of individuals. There are no *de re* modal facts.

In addition, if we went along with Buras, there would be *de re* modal facts, despite the inconstancy of the essentialist claims. However, as Beebee and MacBride note, "[…] what the inconstancy of modal language forces upon us is the recognition that the counterpart relation *is* indeterminate—otherwise the inconstancy of our language would betoken nothing but its failure to be rule-governed" (Beebee & MacBride 2015: 227).

The inconstancy of the truth-conditions for essentialist claims therefore mirrors the fact that there are no counterparts that metaphysically determine which properties deserve to be characterized as the essential properties of individuals; that is, the inconstancy of modal language reflects the fact that there are no *de re* modal facts. Thus, if we recognize the former, we should also recognize the latter, because they come together.

## 8. Conclusion

I have argued that, even though Buras is right in saying that counterpart theoretic ontology can provide metaphysically relevant relations of similarity, this is not sufficient for counterpart theory to accept that individuals have real essential properties.

I have provided two kinds of objections against Buras's thesis.

The first was semantic in character. I have shown that the truth-conditions for essentialist sentences types are inconstant and influenced, sometimes, by considerations about how individuals are represented, even though natural properties are taken into account.

The second was metaphysical in character. I have argued that the acknowledgement of natural properties does not imply the existence of facts of the matter about individuals' essential properties.

In conclusion, I think that counterpart theorists can continue denying that individuals have real essential properties, even if they accept natural properties. Contrary to what Buras claims, the acknowledgment of natural properties in counterpart theoretic ontology does not affect the theory with regard to realist essentialism. If counterpart theory was antiessentialist before natural properties were taken into account, it remains so afterwards.[12]

### References

Beebee, H. and MacBride, F. 2015, "De Re Modality, Essentialism, and Lewis's Humeanism", in Loewer, B. and J. Schaffer (eds.), *A Companion to Lewis*, Oxford: Wiley-Blackwell, 220-37.

Borghini, A. 2005, "Counterpart Theory Vindicated: A Reply to Merricks", *Dialectica*, 59, 67-73.

[12] I would like to thank John Divers and Giorgio Lando for their precious help in improving this paper. I am thankful to Alberto Voltolini for his useful comments and remarks. I am also very grateful to the members of the "Mind, Language and Cognition" curriculum of the FINO PhD program with whom I discussed an early draft of this paper. Finally, I would like to thank two anonymous referees from *Argumenta* for their helpful suggestions.

Borghini, A. and Lando, G. 2011, "Natural Properties, Supervenience, and Mereology", *Humana.Mente*, 19, 79-104.

Buras, T. 2006, "Counterpart Theory, Natural Properties and Essentialism", *Journal of Philosophy*, 103, 27-42.

Della Rocca, M. 1996, "Recent Work in Essentialism, Parts 1 & 2", *Philosophical Books*, 37, 1-13 and 81-9.

Fine, K. 1994, "Essence and Modality", *Philosophical Perspectives*, 8, 1-16.

Guigon, G. 2014, "Overall Similarity, Natural Properties, and Paraphrases", *Philosophical Studies*, 167, 387-99.

Lewis, D. 1968, "Counterpart Theory and Quantified Modal Logic", *Journal of Philosophy*, 65, 113-26.

Lewis, D. 1971, "Counterparts of Persons and their Bodies", *Journal of Philosophy*, 68, 203-11.

Lewis, D. 1979, "Scorekeeping in a Language Game", *Journal of Philosophical Logic*, 8, 339-59.

Lewis, D. 1980, "Index, Context, and Content", in Kanger, S. and Öhman, S. (eds.), *Philosophy and Grammar*, Dordrecht: Reidel, 79-100.

Lewis, D. 1983, "New Work for a Theory of Universals", *Australasian Journal of Philosophy*, 61, 343-77.

Lewis, D. 1986, *On the Plurality of Worlds*, Oxford: Blackwell.

Lewis, D. 2001, 'Truthmaking and Difference-Making", *Noûs*, 35, 602-15.

Marcus, R.B. 1967, "Essentialism in Modal Logic", *Noûs*, 1, 91-96.

Merricks, T. 1999, "Composition as Identity, Mereological Essentialism, and Counterpart Theory", *Australasian Journal of Philosophy*, 77, 192-95.

Nolan, D. 2005, *David Lewis*, Chesham: Acumen.

Quine, W.V.O. 1953a, "Reference and Modality", in Quine, *From a Logical Point of View*, Cambridge (MA): Harvard University Press, 1953, 139-59.

Quine, W.V.O. 1953b, "Three Grades of Modal Involvement", in Quine, *The Ways of Paradox and Other Essays*, New York: Random House, 1966, 158-76.

Sider, S. 1995, "Sparseness, Immanence and Naturalness", *Noûs*, 29, 360-77.

Taylor, B. 1993, "On Natural Properties in Metaphysics", *Mind*, 102, 81-100.

# Propositions as Truthmaker Conditions

*Mark Jago*

*University of Nottingham*

*Abstract*

Propositions are often aligned with truth-conditions. The view is mistaken, since propositions discriminate where truth conditions do not. Propositions are hyperintensional: they are sensitive to necessarily equivalent differences. I investigate an alternative view on which propositions are truthmaker conditions, understood as sets of possible truthmakers. This requires making metaphysical sense of merely possible states of affairs. The theory that emerges illuminates the semantic phenomena of samesaying, subject matter, and aboutness.

*Keywords*: Propositions, Truthmakers, Hyperintensionality, Content, Meaning.

## 1. Introduction

The business of a proposition is to be true or false, depending on how things are. To every proposition corresponds a *truth condition*, displaying how things must be for that proposition's truth. It is natural to take a proposition and its truth condition to be one and the same entity, for that proposition is, by its very nature, true in just those situations set out by its truth condition.

As natural as it is, the view cannot be right, for propositions discriminate where truth conditions do not. A truth condition (as commonly understood) is blind to necessarily equivalent distinctions. Not so for propositions. As the heptasyllabic-happy jargon has it, propositions are *hyperintensional*. A proposition distinguishes between necessarily equivalent situations when they ground its truth in different ways. A proposition's identity goes with the different ways of its being true, and not merely with the different situations in which it is true. Propositions are not truth conditions; they are truth*making* conditions.

Propositions as truthmaker conditions: that is the view I shall articulate and defend. Along the way, I shall ask, just what is a *condition*, and why are propositions not truth conditions? (§2) What does it mean to say that a proposition is a truthmaker condition? (§3) And how can such a view be metaphysically respectable? (§4, §5) The theory that emerges illuminates the phenomena of speakers *saying the same* as one another, but in different ways (§6), and of a statement's *subject matter* and what is about (§7).

*Argumenta* 2, 2 (2017): 293-308
ISSN 2465-2334


DOI 10.23811/47.arg2017.jag

## 2. Truth Conditions and Hyperintensionality

It is common to identify propositions with truth conditions. In classical logic, a truth condition is also a falsity condition. Those who treat propositions as entities need to explain what kind of entity they mean by a *condition*. Suppose we are interested in whether something meets condition *X* in such-and-such situations, and suppose we are interested only in getting a yes-no answer for each such situation. We naturally treat that condition as a function from the situations to the answers, *yes* or *no*. Mathematically, this is a *characteristic function*, and each such function defines a set, containing all and only the input entities for which the function's output is *yes*. It is then both very natural and mathematically elegant to identify the condition itself with that set of situations.

In the case of a truth condition, the input situations are possible worlds and the outputs are *true* or *false*. So we identify a truth condition with a set of possible worlds. Thus, identifying propositions with truth conditions leads us to the view that propositions are sets of possible worlds, defended by Lewis (1986) and Stalnaker (1984; 1976a; 1976b). As a consequence, necessarily equivalent propositions are identical. In particular, there is just one necessarily true proposition (the set of all possible worlds) and one necessarily false proposition (the empty set).

Propositions are not sets of possible worlds, however. One may express necessary truths that are clearly distinct: *that 1=1*, and *that Fermat's Last Theorem is true*, for example. It is not merely that these are distinct sentences. The point is that what they express—what we *say* in uttering them—is so very different in each case.

There are many more ways to make this point. One is via belief: one may believe that 1=1 without believing that Fermat's Last Theorem is true. But I want to avoid this argument from belief, since the identification of propositions as objects of belief is a messy and troublesome business. In saying that David Jones changed the world, I thereby say that David Bowie did, 'Bowie' being the name Jones adopted. But I need not know that, and so need not know (or even believe) that these utterances say the same. I prefer to treat knowledge and belief in a different way (Jago 2014), and to set those concepts aside for present purposes.

Here is another way (not involving belief) of making the distinctness point. The proposition

(1) Lenny is either sleeping or he is not

is about Lenny. By contrast, the proposition

(2) Bertie is either adorable or he is not.

is about Bertie (and not Lenny). *Aboutness* is a relation. So each proposition stands in a relation in which the other does not, and so they are distinct. Yet they are both logical truths; and so they cannot be sets of possible worlds.

Here is yet another way to make the point. Given that Lenny is sleeping, (1) is made true by that very state of affairs. And given that Bertie is adorable, (2) is made true by that, distinct, state of affairs. These propositions stand in the *truthmaking* relation to different states of affairs, and so are distinct entities. So again, they are not sets of possible worlds.

These propositions are more discriminating than truth conditions. A given proposition's nature is not merely to be true or false at a given world, but rather to be made true or made false by specific ways things could be. The way *Lenny*

is, is what makes (1) true, even though it must be true. (Or, more carefully, it is made true by the way Lenny is, if he exists, or by his non-existence, if he does not.) Similarly for Bertie and (2). Propositions are not truth conditions: they are truthmaker conditions.

## 3. Propositions as Truthmaker Conditions

A *truthmaker condition* is a function from possible entities to *yes* or *no*. A *yes* answer indicates that that entity is a truthmaker (or would be, were it to exist) for the proposition in question. Or, more simply, we can identify this characteristic function with the set it defines, so that truthmaker conditions are sets of possible entities. (Typically, in speaking of truthmakers, I will talk in terms of states of affairs. But I do not want to restrict truthmakers to states of affairs, since any entity $x$ whatsoever is a truthmaker for ⟨$x$ exists⟩.) So, as a first pass, propositions are sets of possible entities (of any kind), and we think of those entities as all the possible truthmakers for that proposition.

We do not want to identify a proposition with the set of its actual truthmakers only. We want the proposition to be a condition on what would make it true, were that entity to exist. So the entities in question will have to include merely possible, as well as actual, entities (just as the sets-of-possible-worlds approach to propositions appeals to merely possible worlds). Just how to make sense of this thought is not at all straightforward. Propositions actually exist. They are sets, and so their members actually exist too. But by definition, merely possible entities do not actually exist. This is a deep metaphysical problem for the approach. I will discuss it in detail in §5.

If propositions are sets of their possible truthmakers, then each of those truthmakers must be a single entity. But propositions can be made true by pluralities. ⟨There are pugs⟩ is made true by each individual pug, but also by pairs of pugs, triples of pugs, and, quite generally, by pug pluralities of any size. Treating ⟨there are pugs⟩ as the set of all possible pugs will ignore these pluralities. We might try to avoid this worry by counting all *subsets* of the proposition in question (as well as all of its individual members) among its truthmakers. This approach will identify ⟨there are at least two pugs⟩ with the smallest set with all possible pug-pairs, pug triples, and so on, as its subsets. That is none other than the set of all possible pugs. So this approach incorrectly identifies ⟨there are at least two pugs⟩ with ⟨there is at least one pug⟩.

A better approach is to capture pluralities through their mereological sum. ⟨There are pugs⟩ is the set of all possible pugs and all possible pug-sums. ⟨There are at least two pugs⟩ is the set of all possible two-or-more-pug-sums. Each member of a proposition is then a *full* (possible) truthmaker for that proposition.

A set of possible truthmakers (so understood) is a truthmaker condition. We might want propositions to encode information about their possible falsemakers also. So understood, propositions are truth-and-falsity-maker conditions. We can identify each of these with a set of possible truthmakers plus a set of possible falsemakers. Let us use the notation $|A|^+$ and $|A|^-$ for these sets, respectively. Call the former single-set notion a *single proposition* and the latter double-set notion a *double proposition*. Then both $|A|^+$ and $|A|^-$ are single propositions, and each double proposition ⟨$A$⟩ is a pair of single propositions, ($|A|^+$, $|A|^-$). One very nice feature of double propositions is that, if ⟨$A$⟩ is the pair

($|A|^+$, $|A|^-$), then $\langle\neg A\rangle$ is the pair ($|A|^-$, $|A|^+$). This is because $|\neg A|^+ = |A|^-$ and $|\neg A|^- = |A|^+$.

Double propositions are a good way to distinguish between necessarily false propositions. When $\langle A\rangle$ and $\langle B\rangle$ are distinct propositions, we want to distinguish between $\langle A \vee \neg A\rangle$ and $\langle B \vee \neg B\rangle$. (The inability to do this was part of the criticism of sets-of-worlds account in §2.) But by the same token, we should also distinguish between the necessarily false propositions $\langle A \wedge \neg A\rangle$ and $\langle B \wedge \neg B\rangle$. We cannot do this by identifying propositions with sets of possible truthmakers. (So doing would identify both with the empty set.) Double propositions are a neat solution, for $\langle A \wedge \neg A\rangle$ and $\langle B \wedge \neg B\rangle$ differ in their possible falsemakers. A falsemaker for $\langle A \wedge \neg A\rangle$ is whatever truthmakes either $\langle A\rangle$ or $\langle\neg A\rangle$ (or both) (Fine and Jago 2017). In general, such entities truthmake neither $\langle B\rangle$ nor $\langle\neg B\rangle$. So, although $|A \wedge \neg A|^+$ coincides with $|B \wedge \neg B|^+$, in general $|A \wedge \neg A|^-$ will differ from $|B \wedge \neg B|^-$.

Not every set counts as a single proposition, and not every pair of sets counts as a double proposition. A single proposition $\langle A\rangle$ must be *downwards closed* with respect to grounding. If $x \in \langle A\rangle$ and $y$ is a possible ground for $x$, then $y \in \langle A\rangle$ too. Two points of clarification are in order. First, I am using 'ground' here to mean *full ground*, as opposed to *partial ground*. To illustrate: a conjunction is fully grounded by its conjuncts taken together, and is partially grounded by each of them individually. Second, the closure condition just given makes use of the dyadic notion of grounding: a single entity $y$ as a possible ground for $x$. As above, pluralities of partial grounds, say $x_1$ and $x_2$, are represented as their mereological sum, $x_1 \sqcup x_2$. So we can have $x_1 \sqcup x_2 \in \langle A\rangle$ without $x_1 \in \langle A\rangle$ or $x_2 \in \langle A\rangle$.

We may require that single propositions be upwards-closed with respect to mereological summation: if $x, y \in \langle A\rangle$ then their sum, $x \sqcup y \in \langle A\rangle$. If we do that, then we commit ourselves to impossible entities. If $\langle A\rangle$ has a possible truthmaker $x$ and $\langle\neg A\rangle$ a possible truthmaker $y$, then $\langle A \vee \neg A\rangle$ contains both $x$ and $y$ and so, by the sum closure condition, also contain $x \sqcup y$. But $x \sqcup y$ is a truthmaker for $\langle A \vee \neg A\rangle$! This is an entity that cannot possibly exist. It is a sum of incompatible entities, from different possible worlds. If we want to avoid commitment to such entities, we should restrict the sum closure principle to possible entities: if $x, y \in \langle A\rangle$ and $x \sqcup y$ possibly exists, then $x \sqcup y \in \langle A\rangle$. (In §4, however, I will offer a reason for wanting impossible entities in the theory.) We may also want to ensure that single propositions are *convex*: if $x, z \in \langle A\rangle$ and some part $y$ of $z$ has $x$ as a part (that is, $x \sqsubseteq y \sqsubseteq z$) then $y \in \langle A\rangle$. (Fine (2014b) discusses convexity in relation to content; Fine and Jago (2017) discuss convexity in the context of truthmaker semantics.)

If a set satisfies these conditions, then it counts as a single proposition. That allows many, many arbitrary sets to count as propositions. Take the closure of an arbitrary set, $\{x_1, \ldots, x_n\}$, under the conditions just listed. This is the proposition that $x_1$, or ..., or $x_n$ exists. If the set is upwards-closed under mereological summation, this disjunction can be made true in virtue of any of its disjuncts, or any combination of them, being made true. But there may be additional ways to characterise this set. In general, if (the closure of) $x_1, \ldots, x_n$ are all the possible truthmakers for $A$, then (the closure of) that set is the proposition *that A*. So each proposition is identified with the proposition asserting that at least one of its truthmakers exists, and hence with the proposition that it is made true.

These conditions apply to single propositions, and they apply equally to each component, $|A|^+$ and $|A|^-$, of a double proposition. In addition, we had better rule out any possible entity being in both sets of a double proposition. If some possible $x$ were a member of both $|A|^+$ and $|A|^-$, then it would be possible for both $\langle A\rangle$ and $\langle \neg A\rangle$ to be true simultaneously. But this is not possible, so no possible entity can be in the overlap of $|A|^+$ and $|A|^-$. ($|A|^+$ and $|A|^-$ may overlap only if we accept impossible entities.)

## 4. Metaphysical Worries

There is a serious metaphysical worry facing double propositions. We might identify the double proposition $\langle A\rangle$ with an ordered pair, ($|A|^+$, $|A|^-$). But we might instead identify it with ($|A|^-$, $|A|^+$). Which identification is correct? For the purposes of semantics, either approach is fine. But my interest here is predominantly in the metaphysics of propositions. If we want to know what propositions are, metaphysically speaking, then one choice is right, one wrong; but there is no way to say which.

(If we further identify ordered pairs with sets, we face an additional issue. In general, we can code the pair $(x, y)$ as $\{\{x\}, \{x, y\}\}$, or as $\{\{b\}, \{a, b\}\}$, or as $\{a, \{a, b\}\}$, $\{b, \{a, b\}\}$, or as $\{\{0, a\}, \{1, b\}\}$. Why think that one way gets the nature of propositions right, rather than the others?)

If you do not see a problem here, try this. Consider the pair, ({*that Bertie is snuffling*}, {*that Bertie is not snuffling*}). Assume (for the moment) that this is a proposition. Is it *that Bertie is snuffling* or *that Bertie is not snuffling*? Why? There cannot be any intrinsic differences in the composition of those sets to mark the difference, for the negation of a proposition $\langle A\rangle$ consists in those very same sets, $|A|^+$ and $|A|^-$, but with the order switched: $|\neg A|^+ = |A|^-$ and $|\neg A|^- = |A|^+$. So it seems we need to stipulate which set in the pair comes first, the truthmakers or the falsemakers. Yet there is nothing in the nature of propositions, or the nature of *truth*, which dictates any priority between truth and falsity. The problem is insoluble.

If we cannot make metaphysical sense of double propositions, then we will have to make do with single propositions. But then we must face again the issue of distinct but necessarily false propositions, raised in §3. How should we distinguish them, given that they have no possible truthmakers? We must drop the restriction to possible entities, by allowing propositions to include states of affairs which could not possibly obtain.

Above, we met one way to have impossible entities in our ontology. If possible states of affairs *that A* and *that* $\neg A$ exist, then so does their mereological sum. I take sums of states of affairs to be conjunctive states of affairs: in this case, the (necessarily non-obtaining) state of affairs *that* $A \wedge A$. This state of affairs *that* $A \wedge A$ is distinct from *that* $B \wedge B$ whenever *that A* and *that B* are distinct states of affairs. And that in turn is enough to distinguish $\langle A \wedge A\rangle$ from $\langle B \wedge B\rangle$.

Other impossible cases are not explained so easily. Take the necessarily false proposition $\langle 1 = 2\rangle$. One might think that the very identities of those numbers, 1 and 2, is what makes this false. But on the single-proposition approach, we are limited to possible and impossible truthmakers. What would an impossible truthmaker for $\langle 1 = 2\rangle$ look like?

One suggestion is this: the (necessarily non-obtaining) state of affairs *that {1, 2} is a singleton*. For, were {1, 2} a singleton, 1 would be identical to 2. But this suggestion gets things the wrong way around. The identities of its members make a set the set it is. It is not the properties of the set that fix the identities of its members. Another approach is to take the impossible truthmakers for ⟨1 = 2⟩ to be a state of affairs universally quantifying over properties: *that, for any F, F1 iff F2.* But again, this gets the explanation the wrong way around. It is not that $a = b$ because $a$ and $b$ share all their properties; rather, any property of $a$ is a property of $b$ because $a = b$.

A better approach is available for those who take mathematical entities to be identical to points in a structure. Then, the identity of 1 and 2 is given by relational, structural facts. The (necessarily non-obtaining) state of affairs *that 1 = 2* would be a conjunction of structural facts, identifying the 1-role with the 2-role. Just how this is done (and whether it is plausible) will depend on the details of one's particular structuralist theory.

If there are necessarily existing primitive entities, whose identities are not metaphysically analysable or grounded in more basic facts, then this kind of approach will not cover all cases. We will then be forced to admit some strange ontological ideas. Perhaps there is an *identity* relation, so that *that 1=2* involves the (impossible) instantiation of *identity* with 1 and 2. That is an ugly solution, since for most everyday metaphysical purposes, no *identity* relation is required. Facts of identity are given by the identical entities themselves (which is to say, by each and every thing).

The double propositions account has a much more elegant solution to offer. It treats ⟨1 = 2⟩ as the empty set (since nothing could make 1 = 2) paired with {1 ⊔ 2} (since 1 and 2 together make it the case that 1 ≠ 2). So each account on offer—double propositions, or single propositions with impossible states of affairs—has its benefits and its drawbacks. The former requires us to stipulate, in what would seem an *ad hoc* way, which set in each pair is to count as the truthmakers, which the falsemakers. The latter will probably require the introduction of some dubious ontology. Such is the way in metaphysics. I will put my money on single propositions (with impossible states of affairs).

Both approaches face a further difficulty with propositions such as:

(3) ⟨Propositions exist⟩

(4) ⟨Sets exist⟩.

One might expect the truthmakers for (3) to be all propositions, and truthmakers for (4) to be all sets. Indeed, that result falls out of a general principle: existential truths are made true by the truthmakers for their instances. But this is incompatible with (3) and (4) themselves being sets. Since (3) is a proposition, it would contain itself, contrary to the axiom of regularity (which rules out circular membership chains, $x \in \ldots \in x$. Similarly, if (4) is a set, then it would contain itself; but it cannot.

One may respond that some versions of set theory—non-well-founded theories—allow sets to contain themselves as members (Aczel 1988). I am not tempted by that route. For one thing, I am not sure we can make metaphysical sense of non-well-founded sets, given that sets are grounded by their members. For another, our theory of propositions should not dictate what fundamental mathematical theories should look like.

Even if we set these worries to one side, (4) is simply too big to be a set. If it contained all its truthmakers, it would be the set of all sets. But on pain of contradiction, there can be no such set. There is a similar worry for (3). For each entity $x$, there exists the proposition $\langle x \text{ exists} \rangle$. The possible truthmakers for this are $x$ itself (plus $x$'s grounds). But (3) purports to contain all such sets, and hence purports to be at least as large as the set of all sets. There cannot be such a set.

We might respond with a theory that accepts proper classes, bigger than any set. But then we face the issue: how do we assert the existence of such classes? We are assuming that the proposition $\langle x \text{ exists} \rangle$ is a set-or-class with $x$ as a member. But proper classes are, by definition, members of no set-or-class. So if $x$ is a proper class, then there is no proposition (qua set-or-class) asserting its existence.

These issues run deep. But they are a problem for everyone (who believes in sets). Even if you think there is no such thing as propositions, you still need to explain how the sentences

(5) There are sets
(6) All sets have Ø as a subset

get to be true. These truths require a domain of quantification, which contains all the entities quantified over by those truths. But, on the face of it, both sentences quantify over all sets. That would imply a domain of quantification—a set—containing all sets, which is impossible. (If you want to escape by taking the domain of quantification to be a proper class, just change 'set' to 'class' in the examples to re-introduce the problem.)

Somehow, we meaningfully talk about sets using 'all sets' without thereby including all sets in the domain of quantification. The domain of 'all sets' cannot include the domain of quantification itself. Similarly, the domain of 'some set' cannot include the domain of quantification itself. That seems to be a fact about how the quantifiers work. Their semantics allows 'all sets' and 'some set' to range over all sets except the set specifying that very range.

I propose that the same goes for the quantifiers in (3) and (4): they range over all sets, except the very sets specifying those ranges. Those range-specifying sets are precisely (3) and (4), respectively. So neither (3) nor (4) quantifies over itself, and hence neither is a truthmaker for itself. Both are genuine propositions, on this account. This avoids both the self-membership and the cardinality worry. And, importantly, the result is a consequence of the general semantics for the quantifier 'there are *F*s'. This is a piece of the theoretical jigsaw put in place prior to the account of what propositions are. It does not require us to fiddle with our theory of propositions.

## 5. What are Merely Possible States of Affairs?

I have claimed that propositions are sets, or pairs of sets, of possible (and perhaps impossible) entities. Typically, these entities are states of affairs. On pain of contradiction, not all of the states of affairs thereby quantified over can obtain. But what on earth is a state of affairs that does not obtain? Here are three potential options.

OPTION 1: There exist merely possible concrete states of affairs, making up other possible worlds. 'Obtaining' (relative to world $w$) means ex-

isting at (as part of) world *w*. Non-obtaining states of affairs (relative to our world) are otherworldly states of affairs.

OPTION 2: Some states of affairs do not exist (but remain legitimate objects of quantification). The obtaining states of affairs are those that exist.

OPTION 3: There exist 'ersatz' states of affairs, in addition to the concrete ones. An ersatz state of affairs obtains when it corresponds to some concrete state of affairs.

These approaches are modelled on the main options in the metaphysics of possible worlds. The first takes its cue from the genuine modal realism of Lewis (1986), McDaniel (2004), and Yagisawa (2010). On this approach, all possible worlds are ontologically on a par with our own. The second is a broadly Meinongian approach, defended (in the case of worlds) by Priest (2005). The third approach is based on ersatz modal realism (Adams 1974; Stalnaker 1976a), on which possible worlds other than our own are actually existing ersatz representations.

The genuine realist view of possible states of affairs is a non-starter (even for those who accept entities beyond those that actually exist). I am not wearing a hat, but I could have been. Both states of affairs, *that I am wearing a hat* and *that I am not wearing a hat*, are possible. According to genuine realism about possible states of affairs, reality includes both states of affairs, *that I am wearing a hat* and *that I am not wearing a hat*. Reality is inconsistent! And since the existence of a state of affairs makes the corresponding proposition true, the contradictory propositions ⟨I am wearing a hat⟩ and ⟨I am not wearing a hat⟩ would both be true.

One may respond that those possible states of affairs are parts of distinct possible worlds. No possible world contains both of them (because, although they're each possible, they are not jointly possible). What is possible is whatever obtains at some possible world. So the contradiction, I am wearing a hat and not wearing a hat, is not possible. Consistency is restored.

This response is no solution. A genuine realist (either about possible worlds or about possible states of affairs) needs some standpoint from which she can assert her thesis. But there is no possible world which contains all those entities in which she believes. They are distributed across all the possible worlds. So, if we insist strictly that what is possible is whatever obtains at some possible world or other, then genuine realism (either about worlds or about states of affairs) is ruled impossible from the get-go.

Note that the general problem here does not depend on having negative states of affairs in the ontology. Suppose there is no (actual or possible) negative states of affairs at all. Nevertheless, I could be wearing a completely red hat, and I could be wearing a completely green hat. Those possible states of affairs are metaphysically incompatible. If both exist, as genuine realism entails, then reality is impossible. And we cannot be having that.

The second option mooted above is Meinongian in spirit. It allows that some entities do not exist. On this view, it makes sense to talk about and quantify over entities which lack existence. The suggestion is that merely possible states of affairs be placed in this category. To avoid the problems faced by the genuine realist, the Meinongian must allow that some states of affairs do not act as truthmakers. Rather, she will say, only the existing ones make anything true.

For if all states of affairs act as truthmakers, and there are contradictory (but non-existent) states of affairs, then there are true contradictions (simpliciter), and we are back to the problems from above. So the Meinongian must say that an entity is a truthmaker only if it exists.

But then, what makes it true that some states of affairs do not exist? The only candidate truthmakers for

(7) ⟨Some states of affairs do not exist⟩

are states of affairs that do not exist. But we have just debarred all such states of affairs from acting as truthmakers. So (7) has no truthmakers. It is false. This entails that all states of affairs exist, contrary to the Meinongian view. Meinongianism about possible states of affairs is a non-starter.

Ersatz states of affairs avoid these worries. They count as states of affairs just as rubber ducks count as ducks, which is to say, not at all. They themselves do not constitute something's being the case. They merely *represent* real states of affairs. So they do not make propositions true (other than propositions about the existence of ersatz states of affairs).

What is an ersatz state of affairs, metaphysically speaking? The simplest approach identifies the ersatz state of affairs *that Fa* with the ordered pair containing *F* and *a* themselves, in that order: (*F*, *a*). Such entities look very similar to Russellian structured propositions (King 1995; 1996; Salmon 1986; 2005; Soames 1987; 2008). (They are identical, if we interpret Russellian propositions as set-theoretic tuples.)

As a consequence, the ersatz-truthmaker and Russellian approaches (almost) agree on what *singular* propositions are. For the Russellian, the singular proposition *that a is F* is a structured entity (perhaps a tuple) containing *F* and *a* themselves, in that order. On the ersatz-truthmaker approach, it is the singleton whose sole member is the ersatz state of affairs *that a is F*: $\{(F, a)\}$. But they differ radically on logically complex propositions. On the Russellian approach, a conjunctive proposition $\langle A \wedge B\rangle$ contains both conjuncts and the semantic value of '$\wedge$'. On the truthmaker approach, by contrast, it is the set $\{x \sqcup y \mid x \in \langle A\rangle, y \in \langle B\rangle\}$ of summed truthmakers for each conjunct.

The Russellian and ersatz-truthmaker approach share a common problem. The proposition *that Fa*, by its very nature, represents that *Fa*. But a mere list, tuple, or other structure consisting of *F* and *a* does not, by its very nature, represent that *Fa*. We may interpret some such structure as doing that, as we do for the sentence '*a* is *F*'. Any interpretation given by us is contingent. We could have interpreted the structure some other way, or not at all. So that *F*-and-*a*-involving structure could have represented some other situation, or none at all. The Russellian must say that her proposition *that Fa* might have been some other proposition, or no proposition at all. Similarly, the ersatz state of affairs *that Fa* might have been some other ersatz state of affairs, or none at all, and so its singleton might have been some other proposition, or none at all. But propositions are not like that. Each is *essentially* the proposition representing whatever it represents. So neither the Russellian nor the ersatz-truthmakers account will do.

An adequate solution to our problem should be 'ersatz', in the sense that the entities standing for merely possible states of affairs cannot be genuine states of affairs. But they cannot be 'mere' representations of states of affairs, for these will not maintain the essential link we require between a proposition and what it represents.

My suggestion is this. States of affairs have natures, or essences, just as entities belonging to other categories do, and these natures provide us with a way of talking meaningfully about states of affairs that do not obtain. Here is not the place to argue for the general claim that individuals and properties have natures. But suppose they do and suppose, with Mackie (2006), that such natures are each sufficient for being a given thing (so that, necessarily, *x*'s nature is *y*'s nature only if $x = y$). Suppose further, with Plantinga (1974) that those natures exist necessarily, so that Socrates' nature exists even if Socrates does not. Then, I claim, we have the means to make sense of merely possible states of affairs. These assumptions are substantial commitments to take on. But without them, I cannot see how to understand merely possible states of affairs.

How should we understand the natures of states of affairs, given these assumptions? That will depend largely on our preferred account of states of affairs. Here is one tentative suggestion, built on an Armstrong-style *fundamental tie* view (Armstrong 1997; 2004). On that view, states of affairs have constituents, tied together to form a unified whole. (What about negative states of affairs? I refer the reader to Barker and Jago 2012.) The identities of these states of affairs are given by the identities of their constituents. The nature of *that a is F* is to be the positive state of affairs involving *a*'s possessing *F*. In general, the nature of a state of affairs involves the nature of its constituents. My suggestion is that these natures are unified, structured wholes, just as the corresponding states of affairs are. The nature of *that a is F*, on this approach, involves the natures of *a* and of *F*, bound together by the nature of the fundamental tie.

If the natures of *a* and *F* are necessary existents, then the nature of *that a is F* will be too. So the nature of *that a is F* will exist regardless of whether *a* is *F* (and indeed, regardless of whether *a* exists and whether anything is *F*). For these states-of-affairs-natures are not themselves states of affairs (just as the nature of a given person is not itself a person). So the nature of *that a is F* does not make it the case that *a* is *F*. That is why it is consistent for that nature to exist, even if *a* is not *F*. *a*'s being *F* requires the concrete state of affairs *that a is F* to exist, which is typically a contingent matter. We can then understand 'non-obtaining state of affairs' as picking out a state-of-affairs-nature which corresponds to no actual state of affairs. Let us say that a state-of-affairs-nature is *realised* (at a world) when the corresponding state of affairs exists (at that world).

On this approach, (single) propositions are sets of states-of-affairs-natures. Since these natures actually exist, we have no trouble explaining how propositions actually exist. A (single) proposition is true (at a world) when one of its members is realised (at that world). Importantly, this approach maintains the essential link between a proposition and what it represents, via its would-be truthmakers. A proposition (as a set) is essentially linked to its members, and each of its members (as a state-of-affairs-nature) is essentially linked to a (possible or impossible) state of affairs.

## 6. Same-Saying

We utter declarative sentences to say things to one another. What we thereby communicate is not the utterance itself, since we can say the same thing in different ways. As Frege says:

> If someone wants to say the same today as he expressed yesterday using the word "today", he must replace this word with "yesterday". … The case is the same with words like "here" and "there" (Frege 1956: 296).

Similarly, two people can say the same thing about someone or something in different ways. If I am talking to Anna about her knitting, I will use 'your knitting', she will use 'my knitting', and others might use 'her knitting' or 'Anna's knitting' to say the same thing: *that Anna's knitting is great*.

In these examples, we can contrast what is said with the particular *way* in which it is said. To bring out the idea, suppose Anna and Bob are arguing, Anna insisting that the planet now visible is Hesperus, whereas Bob insists that it is Phosphorus. There is clearly a sense in which they are not really disagreeing at all, for they are both correctly identifying the planet they see. Someone in the know may interject, 'stop arguing, you are saying the same thing!'

Nevertheless, both parties are genuinely informed when they come to learn that the planet is correctly called both 'Hesperus' and 'Phosphorus'. What they lacked was *a posteriori* knowledge, not linguistic competence. This shows that the notion of *what is said* in an utterance does not align with the meaning of the utterance, or with the speaker's beliefs, or with common knowledge in the conversation.

Under what conditions do utterances of two sentences '*A*' and '*B*' say the same thing? (Alternatively, under what conditions do speakers of those utterances say the same thing as one another?) A particularly interesting instance of this question occurs when '*A*' and '*B*' are logically related in a certain way. The question then becomes: which logical operations preserve same-saying? We would like answers to the following kind of question:

(8) Does '$A \vee (B \wedge C)$' say the same as '$(A \vee B) \wedge (A \vee C)$'?
(9) Does '$A \wedge A$' say the same as '$A$'? How about '$A \vee A$'?
(10) Does '$\neg\neg$A' say the same as 'A'?
(11) Does '$\neg$(A $\wedge$ B)' say the same as '$\neg$A $\vee$ $\neg$B'?

Call this general form of question *the logical same-saying issue*. To my knowledge, the issue has not been discussed in the same-saying literature.

The simplest answer to the general same-saying question is this:

(SAMESAYING) *A* says the same as *B* iff $\langle A \rangle = \langle B \rangle$.

Whether that is plausible depends on one's account of propositions. If propositions were sets of possible worlds, it would not be plausible at all. Saying that 1 + 1 = 2 is clearly not the same as saying that properties exist, or that Bertie is either snuffling or not. But all are necessary truths, and hence captured by the same set of possible worlds. Neither would (SAMESAYING) be plausible if propositions were Russellian structured entities. For on that view, 'Bertie is snuffling and wheezing' expresses a distinct proposition from 'Bertie is wheezing and snuffling', and yet these are two ways to say the same thing about Bertie.

I am going to argue that (SAMESAYING) is correct, so long as we understand propositions as truthmaker conditions. This approach provides a plausible general answer to the same-saying question. I will also argue that a truthmaker-based approach provides the only adequate answer to the logical same-saying issue.

If propositions are truthmaker-conditions, then (SAMESAYING) gets the cases involving indexicals and co-referring names right. The possible truthmakers

for 'today is sunny' are defined by taking 'today' fixed in the context of utterance. If today is Thursday 8th September 2016, then the relevant states of affairs capture all the possible ways in which Thursday 8th September 2016 could be sunny. The same goes for 'yesterday was sunny', uttered on Friday 9th. Its possible truthmakers are the same. The two sentences express the same truthmaker condition, and so (SAMESAYING) predicts, correctly, that they say the same thing. The same goes for the 'Hesperus'/'Phosphorus' and the 'my'/'your'/'her' cases.

More interesting is what the truthmaker-condition account says about the logical same-saying issue. It seems clear that distinct but logically related sentences can be used to say the same thing, *in virtue of the logical relation between them*. Any utterance of 'it is warm and sunny' says the same thing as an utterance of 'it is sunny and warm' in the same context. In general, in the same context, utterances of '$A \wedge B$' and '$B \wedge A$' say the same thing. We cannot explain this feature in terms of the necessary equivalence of '$A \wedge B$' and '$B \wedge A$' (or their equivalence in classical logic), because there are equivalent sentences, utterances of which do not say the same thing in a given context. Consider a mathematical example:

(12) I can colour in any map with just three colours, so that no two adjacent areas have the same colour.

(13) I can take one lemon and one orange, and thereby end up with three more fruits than I started.

Both claims are mathematically impossible, and hence (classically) equivalent. Yet utterances of (12) and (13) do not say the same thing. Each speaker claims to be able to do *different* (and, unbeknownst to them, impossible) things. The same holds of logical examples:

(14) The Liar is both true and false.

(15) Claims about large cardinal numbers are neither true nor false.

Here, both statements are classically unsatisfiable (and so classically equivalent), yet they say very different things. Suppose that (14)'s speaker is a dialethist, such as Priest (1979; 1987), who diverges from classical logic in rejecting the explosion principle (that everything follows from a contradiction). And suppose (15)'s speaker is a mathematical intuitionist, such as Dummett (1978; 1993), who rejects excluded middle. It is absurd to think that, in stating their different philosophical positions, they say the same thing as one another. So it is not the case that, in uttering any two classically equivalent sentences, the speakers thereby say the same thing as one another.

A much better account of the logic of same-saying is given by *exact truthmaker equivalence* (Fine and Jago 2017). '$A$' and '$B$' are exactly equivalent when they share all their truthmakers in all truthmaker models. This account predicts that, for each of the following pairs (a/b), utterances of them (in a common context) say the same:

(16a) It is cold and wet.

(16b) It is wet and cold.

(17a) Cath or Dave will turn up, and Ed will turn up.

(17b) Either Cath and Ed will turn up, or else Dave and Ed will.

(18a) Either Cath does not like Dave or she does not like Ed.

(18b) Cath does not like both Dave and Ed.

These pairs are intuitively clear cases of same-saying. So exact truthmaker equivalence looks to be in good standing as an analysis of same-saying.

There are other notions of logical equivalence which treat these cases correctly. *First-degree entailment* (Anderson and Belnap 1963) verifies equivalences (16)-(18), whilst distinguishing between classically equivalent contents. Indeed, relevant logics in general are often seen as ways to preserve content from premises to conclusion in an entailment. (Brady (2006) develops a semantics for a (weak version of) relevant logic in terms of *content inclusion*, for example.) If that is right, then one might expect relevant equivalence to amount to sameness of content, which should in turn amount to same-saying.

But first-degree entailment (and relevant logics in general) do not provide a good account of either same-saying or sameness of content. First-degree entailment treats both $A \wedge (A \vee B)$ and $A \vee (A \wedge B)$ as being equivalent to $A$. But these equivalences do not preserve what is said. Just consider:

(19) Bertie is snuffling, and either he is snuffling or Lenny is sleeping.

(20) Either Bertie is snuffling, or he is snuffling and Lenny is sleeping.

Neither says (just) that Bertie is snuffling, so neither says the same as 'Bertie is snuffling'. So relevant equivalence is not a good criterion for same-saying.

This point is powerful, since just about every logic treats both $A \wedge (A \vee B)$ and $A \vee (A \wedge B)$ as being equivalent to $A$. The truthmaker semantics for exact entailment is one of the few systems that draws semantic distinctions between $A$, on the one hand, and $A \wedge (A \vee B)$ and $A \vee (A \wedge B)$ on the other. So we have a strong argument in favour of analysing same-saying in terms of exact equivalence. Moreover, given the view of propositions as truthmaker-conditions, $\langle A \rangle$ and $\langle B \rangle$ are exactly equivalent iff $\langle A \rangle = \langle B \rangle$. So '$A$' and '$B$' say the same thing (in a context) iff $\langle A \rangle = \langle B \rangle$, just as (SAMESAYING) says.

## 7. Aboutness and Subject Matter

The truthmaker approach to propositions and same-saying also allows for a neat characterisation of a sentence's or proposition's *subject matter*, or what it is about. I will assume we have a fairly good grip on 'being about the same thing'. 'Hesperus'-sentences and 'Phosphorus'-sentences are both about Venus (perhaps amongst other things). We might characterise 'Bertie is snuffling' as being about Bertie and *snuffling*, or we might characterise it as being about whether Bertie is snuffling. (I take these to be distinct but complementary ways of talking about *aboutness*.)

In general, '$A$' and '$B$ can be about precisely the same things and yet not say the same as one another. 'Bertie is snuffling' and 'Bertie is not snuffling' are both about Bertie and *snuffling* (or about whether he is snuffling), yet each says the opposite of the other. Nevertheless, being about the same things is a necessary condition for same-saying:

(ABOUTNESS) $A$ says the same as $B$ only if $A$ and $B$ are about the same thing(s).

As our starting point, let us say that 'Bertie is snuffling' is about whether Bertie is snuffling. We can identify what a sentence or proposition is about with a set of states of affairs. We then define the objects and properties it is about—Bertie

and *snuffling*, in our example—as those that appear as constituents of any of those states of affairs.

There are a number of ways we can implement the first step. The simplest would be to identify the subject matter of a sentence with the proposition it expresses. But this will give us the strange result that $A$ and $\neg A$ have different and indeed incompatible subject matters (since the possible truthmakers for $A$ and $\neg A$ do not overlap). This is the wrong result: $A$ and $\neg A$ are incompatible precisely because they say opposite things about the *same* subject matter.

We improve matters by taking the subject matter of $A$ to be the set of all its possible truthmakers and falsemakers: $|A|^+ \cup |A|^-$. (If we adopt the double proposition account from §3, then we obtain $A$'s subject matter by 'flattening' $\langle A \rangle$ into a single set, $|A|^+ \cup |A|^-$.) This approach gives the correct results for negation: $A$ and $\neg A$ coincide on their subject-matter.

This approach still gives strange results, however. It allows that $A \wedge B$ and $A \vee B$ can have different subject matters. They differ in their truthmakers (and falsemakers) because conjunction pairwise sums together elements from and, whereas disjunction takes their union, $|A|^+ \cup |A|^-$. But this gives incorrect results for subject matter: both are about whatever $A$ is about, plus whatever $B$ is about. They differ in what they say about that subject matter, but not in the subject matter itself.

There are two ways we can avoid this result. One is to take subject matter to be given by all the atomic parts of a sentence's truthmakers and falsemakers:

$$\{x \mid x \sqsubseteq \bigsqcup(|A|^+ \cup |A|^-) \,\&\, x \text{ is atomic}\}.$$

(Here, $\bigsqcup X$ is the sum of all members of set $X$, and 'atomic' means 'having no proper parts'.) The other way is to take subject matter to be the sum of a sentence's truthmakers and falsemakers, $\bigsqcup(|A|^+ \cup |A|^-)$. Both approaches give similar results, given that subject matter (on the second definition) equates to the summed subject matter (on the first definition):

$$\bigsqcup\{x \mid x \sqsubseteq \bigsqcup(|A|^+ \cup |A|^-) \,\&\, x \text{ is atomic}\} = \bigsqcup(|A|^+ \cup |A|^-).$$

One benefit of the second approach is that it allows us to speak of *the* subject matter of a sentence: a unified entity. (On the first approach, by contrast, subject matter is a set, typically containing a plurality. It is somewhat strange, in general, to identify *the* subject matter of a sentence with a set.) The second approach also allows us to make sense of something's literally being a part (as opposed to a member) of a sentence's subject matter.

This approach to subject matter allows us to make sense of notions like *content inclusion* (Fine 2014a, 2014b; Yablo 2014). If we identify $A$'s subject matter with, then $A$'s subject matter includes $B$'s just in case:

$$\bigsqcup(|B|^+ \cup |B|^-) \sqsubseteq \bigsqcup(|A|^+ \cup |A|^-).$$

This notion of inclusion, based on subject matter, ignores whether that subject matter is being affirmed or denied. So, for example, $A \wedge B$'s subject matter will include $A$'s, even though the latter content is incompatible with the former. But we can also define a notion of content inclusion which avoids this consequence. We might say that $A$'s content includes $B$'s content just in case:

$$\bigsqcup|B|^+ \sqsubseteq \bigsqcup|A|^+ \text{ and } \bigsqcup|B|^- \sqsubseteq \bigsqcup|A|^-.$$

On either approach, the notion of content inclusion allows us to analyse locutions like, '*A* is partly about *B*' and '*B* is part of what *A* is about' in terms of *A* including *B*.

Content inclusion (on either approach) does not in general preserve exact truthmaking. $A \wedge B$'s content includes *A*'s content, yet an exact truthmaker for $A \wedge B$ need not exactly truthmake *A*: it may have a *B*-relevant part, which is not relevant to *A*'s truth. (In other words, $A \wedge B$ does not *exactly entail A* (Fine and Jago 2017).) Content inclusion does not even preserve truth. $A \vee B$'s content includes *A*'s content, yet it may be that $A \vee B$ but not *A* is true.

Content inclusion can in turn be used to explain *partial truth*. The intuitive idea is that 'Hilary Putnam was one of the greatest female philosophers' is partly true (since he was one of the greatest philosophers), but not wholly true (since he was not female). A simple take on partial truth has it that *A* is (at least) partly true when it content-includes some (wholly) true *B*. 'Hilary Putnam was one of the greatest female philosophers' content-includes both 'Hilary Putnam was a philosopher' (true) and 'Hilary Putnam was female' (false) and so, on this analysis, is partly (but not wholly) true. (Fine 2016 gives an alternative account in terms of *analytic containment*, based on Angell 1989.)

There is clearly much more to be said about aboutness, subject matter, and the various notions of content inclusion. Yablo (2014) discusses these concepts in detail (he offers a fine-grained possible worlds-based account). My suggestion here is that the truthmaker approach offers a natural and elegant way to account for these concepts.

## 8. Conclusion

Propositions are not truth-conditions; they are truthmaker conditions. Metaphysically, truthmaker conditions are sets of the natures of actual and merely possible entities (typically, but not exclusively, states of affairs). Working with the natures of entities (rather than the entities themselves) allows us to capture the identities of merely possible entities without descending into paradox. Logically, the identity conditions on propositions is given by the logic of strict truthmaker equivalence. And semantically, the theory of propositions as truthmaker conditions illuminates *samesaying*, *subject matter*, and *aboutness*.

### References

Aczel, P. 1988, *Non-Well-Founded Sets*, CSLI Lecture Notes Vol. 14, Stanford: CSLI Publications.

Adams, R. 1974, "Theories of Actuality", *Nous*, 8 (3), 211-31.

Anderson, A. and Belnap, N. 1963, "First Degree Entailments", *Mathematische Annalen*, 149 (4), 302-19.

Angell, R.B. 1989, "Deducibility, Entailment and Analytic Containment", in Norman J. and R. Sylvan (eds.), *Directions in Relevant Logic*, Dordrecht: Kluwer Academic Publishers, 119-43.

Armstrong, D. 1997, *A World of States of Affairs*, Cambridge: Cambridge University Press.

Armstrong, D. 2004, *Truth and Truthmakers*, Cambridge: Cambridge University Press.

Barker, S. and Jago, M. 2012, "Being Positive About Negative Facts", *Philosophy and Phenomenological Research*, 85 (1), 117-38.

Brady, R. 2006, *Universal Logic 2006*, Stanford: CSLI Publications.

Dummett, M. 1978, *Truth and Other Enigmas*, Cambridge (MA): Harvard University Press.

Dummett, M. 1993, *The Seas of Language*, Oxford: Oxford University Press.

Fine, K. 2014a, "A Theory of Truth-Conditional Content I: Conjunction, Disjunction and Negation", *Unpublished manuscript*.

Fine, K. 2014b, "A Theory of Truth-Conditional Content II: Subject-matter, Common Content, Remainder and Ground", *Unpublished manuscript*.

Fine, K. 2016, "Angelic Content", *Journal of Philosophical Logic,* 45 (2), 199-226.

Fine, K. and Jago, M. 2017, "Exact Truthmaker Logic", *Unpublished draft*.

Frege, G. 1956, "The Thought: A Logical Inquiry", *Mind*, 65 (259), 289-311.

Jago, M. 2014, *The Impossible*, Oxford: Oxford University Press.

King, J. 1995, "Structured Propositions and Complex Predicates, *Noûs*, 29 (4), 516-35.

King, J. 1996, "Structured Propositions and Sentence Structure", *Journal of Philosophical Logic*, 25 (5), 495-521.

Lewis, D. 1986, *On the Plurality of Worlds*, Oxford: Blackwell.

Mackie, P. 2006, *How Things Might Have Been: Individuals, Kinds, and Essential Properties*, Oxford: Oxford University Press.

McDaniel, K. 2004, "Modal Realism with Overlap", *Australasian Journal of Philosophy*, 82 (1), 137-52.

Plantinga, A. 1974, *The nature of necessity*, Oxford: Oxford University Press.

Priest, G. 1979, "Logic of Paradox", *Journal of Philosophical Logic*, 8, 219-41.

Priest, G. 1987, *In Contradiction: A Study of the Transconsistent*, Dordrecht: Martinus Nijhoff.

Priest, G. 2005, *Towards Non-Being*, Oxford: Clarendon Press.

Salmon, N. 1986, *Frege's Puzzle*, Cambridge (MA): MIT press.

Salmon, N. 2005, *Metaphysics, Mathematics, and Meaning*, New York: Oxford University Press.

Soames, S. 1987, "Direct reference, Propositional Attitudes and Semantic Content", *Philosophical Topics*, 15 (1), 47-87.

Soames, S. 2008, "Why Propositions Cannot Be Sets of Truth-Supporting Circumstances", *Journal of Philosophical Logic*, 37 (3), 267-76.

Stalnaker, R. 1976a, "Possible Worlds", *Noûs*, 10 (1), 65-75.

Stalnaker, R. 1976b, "Propositions", in MacKay, A. and D. Merrill (eds.), *Issues in the Philosophy of Language*, New Haven: Yale University Press, 79-91.

Stalnaker, R. 1984, *Inquiry*, Cambridge (MA): MIT Press.

Yablo, S. 2014, *Aboutness*, Princeton-Oxford: Princeton University Press.

Yagisawa, T. 2010, *Worlds and Individuals, Possible and Otherwise*, New York: Oxford University Press.

# Husserlian Intentionality and Contingent Universals

*Nicola Spinelli*

*King's College London*
*Hertswood Academy*

*Abstract*

Can one hold both that universals exist in the strongest sense (i.e., neither in language nor in thought, nor in their instances) *and* that they exist contingently—and still make sense? Edmund Husserl thought so. In this paper I present a version of his view regimented in terms of modal logic *cum* possible-world semantics. Crucial to the picture is the distinction between two accessibility relations with different structural properties. These relations are cashed out in terms of two Husserlian notions of imagination: world-bound and free.

After briefly presenting the Husserlian framework—his intentionalism, idealism and how universals figure in them—I set up my modal machinery, model the target view, and show that, depending on the chosen accessibility relation, the necessary or the contingent existence of universals can be derived. Importantly, since for Husserl both relations are bona fide, both derivations are legitimate. In Husserl's philosophy, then, there is room for both necessary and contingent universals.

*Keywords*: Husserl, intentionality, modality, imagination, universals.

Some philosophers believe in universals and some dismiss them as a myth. The former think that, in addition to—say—all red things, there is a further thing: the property of being red. Disbelievers, by contrast, have it that the property of being red is at worst a mere *façon de parler*, at best a linguistic, conceptual or mathematical construction, but certainly not a genuine 'thing'. Interestingly, in *both* camps virtually everyone agrees that if universals exist, they exist as a matter of necessity—or, as disbelievers would put it, that if they existed, they would exist as a matter of necessity.

What if all of them were wrong? That is to say, what if universals, conceived as bona fide objects distinct from, and irreducible to, their instances existed contingently? Does the notion of a contingent but genuine universal even make sense? Few philosophers have maintained that it does. One of them is Edmund Husserl, who in *Experience and Judgement* (1939) holds that universals come in two kinds—pure universals, or *eide*, and empirical universals—only one of which, namely the

*Argumenta* 2, 2 (2017): 309-325
ISSN 2465-2334


DOI 10.23811/48.arg2017.spi

pure, enjoys necessary existence (Husserl 1973, §82). In this paper I present a version of his position, regimented in terms of modal logic and possible-world semantics.

I need to spend at least a few words on the development of Husserl's outlook on universals. While doing so, however, I will not embark in fine-grained scholarly questions (Husserl says so-and-so in book *x* but denies it in manuscript *y* and then takes up an intermediate position in letter *z*). Not because I think they are trivial (they are not), but because here I am interested in the view itself. The adamant Husserlian scholar may thus read this paper as being about a view that is not *Husserl's* but *Husserlian*—as at the very least it does resemble Husserl's own view:

> Empirical generalities [...] bring with them the copositing of an empirical sphere in which they have the place of their possible realization in particulars. If we speak of plants, cities, houses, and so on, we intend therewith in advance *things of the world*, and in fact the world of our actual, real experience (not a merely possible world); accordingly, we think of these concepts as *actual* generalities, that is, bound to this world (Husserl 1973: 330).

Notoriously, at the time of the *Logical Investigations* (1900-1901) Husserl was a Platonist: his view was that universals—items such as the property of being red or the relation of being friends with someone—are non-spatiotemporal objects existing independently of our minds and irreducible to their instances (even to their possible instances). The *Investigations* are indeed an attempt at investigating our epistemic access to universals Platonically understood, as well as to other ideal objects (these days we call them 'abstract') such as numbers and meanings.

By 1913, however, Husserl had become an idealist, and his outlook on universals had changed accordingly: he still retained the view that universals are irreducible to their instances, but he now dropped mind-independence.[1] However, he did not become a nominalist for that. He thought that universals are independent of any particular subject's mind (and thus, in particular, are not 'in' the subject's mind), but that their existence-conditions—along with the existence-conditions of every object, including spatiotemporal ones—should be spelled out in terms of consciousness and, ultimately, of intentionality. Some scholars, I should mention, deny that Husserl ever became an idealist. That Husserl was indeed an idealist is an assumption of this paper; if you need convincing, my suggestion is to look at the case A.D. Smith (2003) makes, which I endorse.

Now, although a number of idealistic existence-conditions for universals can be made out from Husserl's texts, I will only rely on one—which is, in a sense to be specified in due course, prior to (required by) the others. My treatment here may be seen as the core of a wider regimentation including all the conditions. Obviously a comprehensive treatment would be desirable; but first things first. Restricting attention to the relevant, necessary but not sufficient existence-condition simplifies things considerably and yields a reliable and fairly elegant picture, later to be developed on a step-by-step basis.

As far as the problem itself goes—whether, that is, universals may exist contingently—there is of course a disadvantage in working within Husserlian ideal-

[1] According to Mark Van Atten (2007), Husserl also dropped atemporality—albeit in a qualified sense. I tend to agree, but I will disregard the issue in this paper.

ism: the resulting discussion will not map on the current (or most of the traditional) debate on universals, because most believers in universals are realist, and most disbelievers disbelieve what believers believe, i.e., that universals exist in a realist sense (that is, mind-independently).[2] Nonetheless, I think, Husserl's view is worth taking into account for at least three reasons. One historical: seldom if ever have Husserlian scholars touched upon, let alone unpacked, this particular issue—contingent universals (Sowa 2007 is a notable exception). Secondly, Husserlian idealism is a curious environment, a natural habitat to some interesting breeds such as, for example, intuitionistic choice sequences: the only mathematical objects that develop in time (Van Atten 2007, 2015). I see contingent universals as drinking from the same pools and grazing the same grass. The third reason is that, I suspect, once regimented as I propose to do here, Husserl's view can be extended to realistic environments—though this, I have to admit, at this stage is mere conjecture. After briefly introducing, in Section 1, Husserlian intentionality and idealism—not comprehensively, but rather only as much as I need for my purposes—I will proceed to my regimentation of the theory. This will be in terms of basic modal logic and possible-world semantics. In Section 2 I will model the main concepts and claims presented in Section 1 by introducing some bespoke non-logical predicates and formalising the two main sentences to be proved: that universals exist necessarily and that universals exist contingently. In Sections 3 and 4 I will discuss accessibility between possible worlds and imagination. In the Husserlian framework, accessibility is to be cashed out in terms of imagination. Husserl has two notions of imagination: world-bound and free. In Section 3 I will characterise them, while in Section 4 I will show how they yield two distinct accessibility relations and illustrate their structural properties. Finally, in Section 5 I will show that, depending on the chosen type of imagination, the necessary or the contingent existence of universals can be derived.

## 1. Universals in Husserlian Idealism

The core idea of Husserlian idealism is that the conditions for the existence of objects, including universals, are to be specified in terms of consciousness. And since for Husserl conscious mental performances are intentional, in Husserlian idealism an object exists if and only if it meets certain conditions spelled out in terms of intentionality. In this section, then, I will sketch a Husserlian theory of intentionality, expand it into idealism, and show how universals figure in it, especially as regards their existence-conditions. As I mentioned earlier, I will only rely on one among several such conditions, hoping that my treatment may be expanded to include the others.

### 1.1. Husserlian Intentionalism

For Husserl, as well as for other philosophers, a wide class of mental acts are 'intentional': they are, by their own nature, *directed to something*. To think, for example, is to think of something; to love is to love something; to fear is to fear

[2] Clearly, by saying that Husserl was not a realist about universals I do not mean to say that he was an eliminativist: quite the opposite! I just mean that if a realist about universals is one that believes that universals exist mind-independently, then Husserl was not a realist but an idealist. Yet he did not doubt that universals exist. The realism-idealism issue here is, if you will, metaphysical, not ontological.

something; to hallucinate is to hallucinate something; and so on. And it seems that this directedness is part of what those acts are. One way of putting this is as follows: intentional acts link, by their own nature, a subject (a mind) to an object. Intentionality is thus a relation, whose first-place relatum is the subject and whose second-place relatum is the 'intentional object'. As some intentionalists point out, however, the intentionality relation has the following peculiarity: that its second-place relatum, the intentional object, is 'existence-independent' (Smith and McIntyre 1982, Drummond 1990).

Intentionalists of this stripe, including Husserl but excluding notorious intentionality-theorist such as John Searle (Searle 1973), deny that the objects of intentional acts are entities—where an entity is something that exists, that is part of reality. Of course, for there to be an intentional act there must be a subject that performs it (Husserl would say: that 'lives' it). However, the object of the act need not exist. Why think so?[3] Because it is a phenomenological fact about consciousness that we can be aware of non-existents: we can think of Santa Claus, for example, or—if our calculus is rather rusty—look for the unique and completely determined result of $\int x + 1 \, dx$; or we can fall in love with a character in a book, or fear the ghost of Abraham Lincoln (which reportedly haunts the White House)—or, finally, hallucinate a fat man in our empty doorway. Even in these cases, intentionalists hold, we are aware of something, we have something over and against our consciousness, just as well as in cases in which what we are aware of exists in reality.

As A.D. Smith, a prominent intentionalist, puts it:

> Central to intentionalism is the denial that the expression 'is aware of' must express a relation between two entities. On this view, to speak of an object of awareness is not necessarily to speak of an entity that is an object of awareness: for some objects [of awareness] do not exist (Smith 2002: 224).

That being their position, intentionalists need a way to construe the notion of 'object' that does not include the claim that an object, as such, exists. What we may call the Husserlian way is as follows. Talk of a mental performance's having an intentional object is, from an ontological standpoint, *just* talk of *the mental performance*, just the description of a particular way in which the subject is minded. So that, for example, the sentence 'I am thinking of Lincoln's ghost' does not by itself carry a commitment to the existence of a certain object, Lincoln's ghost; it only carries a commitment to the existence of a certain mental act, my thinking of the ghost, whose descriptive character (as Husserl would put it: whose descriptive essence) is best captured in terms of directedness to an object.

It is in this sense that, for the Husserlian intentionalist, the notion of intentional object has absolutely no ontological import (apart from implying that there exists a subject who is minded in a particular way), and is, rather, merely phenomenological: it only serves to adequately describe a certain class of experiences (namely, the intentional ones). And it is in this sense that existence and non-existence do not accrue to intentional objects qua intentional objects. In Husserl's words:

[3] Different answers to the question may be found in the literature. For a discussion, see Spinelli 2016. Here I will pursue what we may call the phenomenological line.

> […] only one thing is present, the intentional experience, whose essential descriptive character is the intention [i.e., the specific directedness] in question. […] If this experience is present, then, *eo ipso* and through its own essence, the intentional 'relation' to an object is achieved and an object is 'intentionally present'; these two phrases mean precisely the same. And of course such an experience may be present in consciousness […] although its object does not exist at all. […] The object is 'meant', i.e., to 'mean' it is an experience, but it [the object] is then merely entertained in thought, and is nothing in reality. […] If, however, the intended object exists, nothing becomes phenomenologically different. […] I think of Jupiter as I think of Bismarck, of the tower of Babel as I think of Cologne Cathedral, of a regular thousand-sided polygon as of a regular thousand-faced solid (Husserl 2001: 98-99).

A helpful way of understanding the view is in terms of reduction. As Smith has it, Husserlian intentionalism is an *ontologically reductive* view of intentional objects, because it holds that intentional objects are nothing over and above the experiences of which they are the objects. It is not, however, a psychologically (or, better, phenomenologically) reductive view: because for the Husserlian intentionalist the only adequate way of describing the experiences in question is as being directed towards objects. That is what Husserl means when he says that directedness belongs to the 'essential descriptive character' of the acts. In the analytic tradition, Smith is perhaps the prominent purveyor of this outlook (see his treatment of hallucination in Smith 2002).

Of course, Husserlian intentionalism is not immune to objections. Here are two. First, why trust phenomenology implicitly, as Husserl does? Could not phenomenology be deceiving? If it is, then what the Husserlian intentionalist takes to be facts are really no facts at all—and the whole theory goes with them. Secondly, does it make sense to speak of a relation with a non-existent relatum? Is it not written into the notion that for a relation to obtain its relata must first exist? If the answer is yes, then the Husserlian intentionalist is in trouble. And the fact that intentionality is a relation between an existent relatum and a non-existent one—rather than a relation between two non-existents—might be even more troubling.

The Husserlian could reply that while the claim that a relation must have relata is uncontroversial, the claim that a relation must have *existent relata* is far from obviously true. If correct, this would defuse the objection. Be that as it may, since my aim here is not to defend Husserlian intentionalism, but rather work within it, I will assume that these and other issues can be addressed satisfactorily, and carry on.

### 1.2. Husserlian Idealism

So we have seen what intentional objects are in Husserlian intentionalism, and in particular that existence and non-existence do not accrue to them as such: some intentional objects exist, some do not. A paradigm example is the pair perception-hallucination: both are intentional performances, but in the perceptual case the object exists, while in the hallucinatory case it does not. This is a distinction that we obviously want to make. Yet how exactly is it to be made?

Someone that we might call an intentional realist (a realist that buys into Husserlian intentionalism) would answer something like: 'That is actually quite easy: even though in both the perceptual and the hallucinatory case there is an intentional object, in the first but not in the second case the object is really out there'. An answer that, for all its *prima facie* plausibility, would not satisfy Husserl.

The reason is that it is simply not clear what 'really out there' means. What is doing the explanatory work (explanatory, that is, with respect to the concept of existence)? Certainly not 'really': appealing to a distinction between appearance and reality will not do, because what the distinction turns on—the fact that in one case the object exists, is part of reality, and in the other it does not—is precisely what needs explaining.

So the important bit must be 'out there'. But what does it mean? Clearly it cannot mean that in perceptual cases, unlike in hallucinatory ones, the object is 'over and above the mental act', because that is just to say that the object exists—and this is, again, what needs explaining. Nor can it point to the fact that in the perceptual but not in the hallucinatory case the object is in space and time: because surely a hallucinated object is hallucinated as being in space and time. Thus, either the claim is that in the perceptual case the object is 'really' in space and time—and we are back to the previous point; or the claim is off-target, because, again, hallucinated objects are in space and time too.

A different approach available to the intentional realist is in terms of mind-independence: in the perceptual case, the object of awareness is not simply an object of awareness, but is mind-independent—it belongs to the mind-independent world—whereas in the hallucinatory case it is only an object of awareness. While it is not obvious that this illuminates the concept of existence in any way, Husserl definitely thinks, and repeatedly states, that the very idea of a mind-independent world is 'nonsense' (quoted in Smith 2003: 182; also Husserl 1969: 204). This cuts the discussion short; yet it is unclear to me whether for him it is a motivation for, equivalent to, or a consequence of idealism.

Finally, the intentional realist might take existence as a primitive: 'At the end of the day', he will say, 'we must have primitives; and what better primitive than existence?'. And here we come to the main reason why Husserl is not satisfied with intentional realism. The reason is that he was a Cartesian: he thought that, outside or at the edge of our 'natural', everyday attitude towards the world, ontological statements are always liable to be doubted. Appealing to a mind-independent world to account for the existence of some intentional objects seems therefore suspicious to him: because, he maintains, while in the natural attitude we simply and firmly believe that there is such a mind-independent world, philosophically that belief needs justification. Hence the need, which since as early as 1906 was the motivation behind the so-called 'transcendental turn', to give every ontological statement a phenomenological reading (this is what the notorious 'epochè' and 'transcendental reduction' amount to).[4] Thus, although we obviously want to say that some intentional objects exist while others do not, the only satisfactory way of cashing this out is, for Husserl, again in terms of consciousness—of intentionality.

Admittedly, this latter form of "methodological" idealism is weaker than the 'mind-independence is nonsense' stance we saw before. It is perhaps something of a stretch even to call it idealism. Yet, insofar as it includes the claim that existence is to be understood in terms of consciousness, I think the qualification is justified.

[4] This implies that phenomenology is the fundamental element of Husserl philosophy. Oddly, this view has been challenged (Smith 2007). I do not believe there is anything to be said for any of the opposing theories. Yet to defend my position would take me too far afield. It is enough, for the sake of integrity, to have raised awareness of a debate lurking in this connection.

Thus, for Husserl, in order to account for the distinction between existing and non-existing intentional objects we should look at the phenomenology of the relevant intentional performances. And what we see once we do so is, for him, that what distinguishes existing from non-existing intentional objects—those perceived, say, from those hallucinated—is that the former are public in a way in which the latter are not.

A hallucinated object will be available for the hallucinating subject but not for others. In fact, if prompted other subjects will deny it exists. Now you need not be an idealist to accept this. The realist can say: of course, if the object does not exist, non-hallucinating subjects will—at least in principle—deny that it does. But they will do so *because* it does not exist, not the other way round! Where is the explanation, then?

What makes Husserl an idealist is that he thinks that intersubjective confirmation is not only a necessary condition for the existence of an intentional object, but also a sufficient one. Indeed, Husserl's phenomenological account of existence consists in providing increasingly comprehensive accounts of what it is for an intentional object to be public in this sense, i.e., for it to be available *as an intentional object* to a community of subjects.

Notice that Husserl's view is not, after all, too far removed from common sense. Consider the following question: 'What do you mean when you say that the Eiffel Tower, this thing you are seeing right now in front of you, really exists?'. It is just natural, I submit, to answer it by saying something like: 'I mean that I am seeing it now, and that if I turn my head and then turn it back again I will still see it; and that if I had been here yesterday I would have seen it—and you too; and that anyone who stands here and is not blind will see it; and that anyone who has stood here since 1889 and was not blind has seen it'. The only difference—and what a difference—is that for Husserl 'I mean' is to be taken seriously, i.e., as an equivalence (in fact, probably as an analysis).

Existence, then, is cashed out in terms of the intentional performances of what Husserl calls 'transcendental intersubjectivity'. In other words, for an intentional object to exist is for it to be available not only to one subject at a given time and under specific circumstances, but to any possible subject at any suitable time and under any suitable circumstances.

Husserl's view is complex and is never put forward entirely in any one particular text. So it would be impossible for me to do it justice here. Smith 2003 is a good place to start. Here I will simply work within the Husserlian framework and without questioning it. This, of course, is not to say that the framework faces no difficulties. For example: given that intentional objects are nothing over and above a subject's mental performances, how can an intentional object—the *same* intentional object—be available to two distinct subjects, or in fact even the same subject at two particular points in time, let alone to an all-comprehensive, actual and possible transcendental intersubjectivity? To this question Husserl does have a (good?) answer, of course, and it is in terms of yet further intentional acts in which intentional objects are identified. Daredevils may approach Husserl 2005 for further detail. See also Hopkins 2016 for a critical discussion.

### 1.3. Universals and Existence

There is evidence that for Husserl the existence of spatiotemporal objects and the existence of ideal objects, such as numbers and indeed universals, are to be accounted for in similar ways. For example the following passage:

> The *transcendence of the world* [i.e., for my purposes, the reality or the existence of the world] … is of the same species as the *transcendence of numbers* and other [ideal] objects (Husserl 1959: 180).

A universal will thus be an existent intentional object, as opposed to a mere intentional object, if it is at the very least such that, in addition to being something someone has actual epistemic access to (e.g., something someone is actually thinking about), is something that some other possible subject has epistemic access to (e.g., thinks about). In other words, if it is public at least in a minimal sense.

In an exhaustive picture, further conditions would have to be put on the existence of universals. For example, the universal in question must not generate contradictions—the underlying thought being that a contradictory universal will be disconfirmed in a similar way as a hallucinatory object is. Thus, although the property of being the set of all sets that are not members of themselves is, under our basic condition, an existing intentional object, since it gives rise to Russell's antinomy it does not exist under the stricter, suggested condition. Here however, for reasons I have already given, I will only work with the basic criterion. There is still some work to do to get even that criterion right, however. At a first approximation the condition looks as follows:

> (ExU1) For a universal *u* to exist in the actual world is (partly) for it to be not only an intentional object in the actual world, but also an intentional object in some possible world accessible from the actual.

Surely, though, if we are prepared to say that a given universal exists, we must also be prepared to concede that it exists *regardless* of whether someone in the actual world is thinking of it or has even ever thought of it. Take for example the property of being a pathological function (i.e., a function that is uniformly continuous in R but nowhere differentiable). That property was never an intentional object before 1830, when Bernard Bolzano discovered its first instance.[5] Surely, however, if the property exists then it existed before 1830. Moreover, it *would have existed* even if no one had ever discovered or thought about pathological functions. Otherwise it is dubious that we would be targeting a plausible, if idealistic, concept of existence—at least as far as universals are concerned. The existence-condition as it stands, then, is too strict. Let us relax it a little:

> (ExU2) For a universal *u* to exist in the actual world is (partly) for it to be an intentional object in some possible world accessible from the actual.

Thus, in Husserlian idealism a universal exists only if, at the very least, it is possible that it should be an intentional object. Moreover, and consequently, it exists necessarily only if it is necessarily possible that it should be an intentional object; and it exists contingently if it does not exist necessarily, i.e., if it is not necessarily possible that it should be an intentional object.

A final remark—almost a side note—to dispel misunderstandings. We have seen that Husserl's view is that the existence conditions of both spatiotemporal

[5] In fact, pathological functions only became mainstream thanks to a 1872 paper by Karl Weierstrass, and that is probably when the concept became established. Incidentally, Husserl, who had studied in Berlin under Weierstrass, was also his assistant for a short time between 1882 and 1883.

and non-spatiotemporal (ideal) objects are to be cashed out in terms of consciousness and, in fact, in similar ways. But, of course, the two accounts will differ at some level. For one thing, different types of intentional acts will be relevant in each case. In the case of spatiotemporal objects, perception will be paramount, while in the case of ideal objects thought will. So, even though ExU1 and ExU2 are, in an abstract sense, applicable to both spatiotemporal and non-spatiotemporal objects (as surely the former, if they are to be real, must be intentional objects not only in the actual world, but also in some possible world sufficiently close to the actual), once the mental performances with respect to which the relevant items are intentional objects (in the actual or in some possible world) are specified the two accounts will differ significantly.

Nonetheless, I wish to stress, they are analogous. This is, if you will, a phenomenological and idealistic take on the Meinongians' distinction between subsistence and existence. Notice, moreover, that ExU1 and ExU2, as well as their analogues for spatiotemporal objects, are only necessary existence conditions: to characterise the two types of existence fully, sufficiency clauses are needed.

### 1.4. Universals and Genetic Basis

I have already stressed that at no stage was Husserl a nominalist. However, he always maintained that without the prior consciousness of similarities among particulars there would be no consciousness of universals. This is not to say that for Husserl consciousness of universals reduces to consciousness of particulars: the second *Logical Investigation* is just a long argument against precisely that view. It is to say, however, that for him intentional performances that have universals as their intentional objects are based or grounded on, or at least made possible by, systems of intentional acts in which the subject becomes aware of the relevant similarities among particulars. Long and very technical analyses of the relations between these two types of acts, or systems thereof, may be found in *Experience and Judgement*.

What matters for my purpose, however, is that in Husserl's view for there to be an act directed at universal $u$ in a given world $w$—and thus for $u$ to be an intentional object in $w$—there must be some $u$-particulars in $w$ such that the recognition, on the part of the subject, of their similarities forms the basis for the $u$-directed act. If the universal in question is the property of being black, for example, for it to be an intentional object in $w$ there must be black particulars in $w$. *Ex post*—that is, once universal $u$ is brought to consciousness—we say that those particulars are a (proper) subset of the extension of $u$: some of its instances. In fact, however, they are a privileged subset: as I will refer to them, they are the *genetic basis* of $u$.

In a world without a suitable genetic basis, $u$ is simply not brought to consciousness—it is not an intentional object. That is not to say that $u$ does not exist in that world: because existence is, as ExU2 states, the possibility of being an intentional object. Hence, for $u$ to exist in a world $w$ with no $u$-particulars in it, it is enough that 1) there is a possible world $w_1$ accessible from $w$ with $u$-particulars in it, and 2) that $u$ is an intentional object in $w_1$.

Some remarks about possible worlds are in order. Hintikka 1975, Hintikka and Harvey 1992 and Smith and McIntyre 1982 have attempted to reconstruct (and to an extent improve) Husserl's theory of intentionality by regimenting it in

terms of possible-world semantics. They have been criticised for that: see e.g. Mohanty 1985, 1999, Drummond 1990, Welton 2000. Now, what I am attempting to do here has nothing or very little in common with those contributions. I need, however, to say something about the notion, common to most critics, that since there is ample room to doubt that Husserl ever took talk of possible worlds seriously (Mohanty 1985: 35-39—but to be fair both Hintikka 1975 and Smith and McIntyre 1982 acknowledge this), it is at least dubious that one can even start to think of legitimately reconstructing any part of Husserl's philosophy in terms of possible worlds.

This line of reasoning is fallacious: you need not take possible worlds seriously to use them. Modal actualists, for example, think that talk of possibilia ultimately reduces to talk of actualia. That, however, does not prevent them from using such talk for semantic purposes: see for example Fine 2005a, 2005b. Interestingly, Wallner 2014 argues that Husserl himself was an actualist of sorts—albeit perhaps implicitly. He calls the Husserlian brand of actualism 'phenomenological actualism'. So modal reasoning can be framed in terms of possible worlds even if possible worlds are not taken seriously from a metaphysical standpoint: one need only be a (tolerant) modal actualist, as Husserl himself arguably was.

## 2. Modal Machinery

Let 'E!($u$) at $w$' mean 'Universal $u$ exists in possible world $w$', and 'Int($u$) at $w$' mean 'Universal $u$ is an intentional object in world $w$'. Let also the following be the necessary and sufficient condition for the existence of $u$ in world $w$ (as I said, it is in fact only necessary; I will treat it as necessary and sufficient for the sake of simplicity, and explain why that is no problem at the very end of the paper):

[Ex] E!($u$) at $w \leftrightarrow \Diamond$Int($u$) at $w$.

Let @ be the actual world. Then, given [Ex] and the usual truth conditions for $\Diamond$:

[Ex@] E!($u$) at @ $\leftrightarrow \exists w^{@\mathrm{A}w}$(Int($u$) at $w$).

Superscripts such as '@A$w$' mean 'world @ has access to world $w$'. They are up there because I need subscripts to indicize worlds.

There are two modalised versions of [E$x$@]. The first gives the formulation of the possible existence of $u$ at @. It follows again from [E$x$] and the truth conditions for $\Diamond$. The only difference is that now we have a double diamond: because possible existence of $u$ at @ is the possibility of the possibility of $u$'s being an intentional object at @. A few calculations:

$\Diamond$E!($u$) at @
$\Diamond\Diamond$Int($u$) at @
$\exists w^{@\mathrm{A}w}(\Diamond$Int($u$) at $w)$

yield:

[$\Diamond$E$x$@] $\Diamond$E!($u$) at @ $\leftrightarrow \exists w^{@\mathrm{A}w}(\exists w^{w\mathrm{A}w1}$(Int($u$) at $w_1$)).

Then there is the formulation of the necessary existence of $u$ at @. It follows from [Ex] and the usual truth conditions for $\Box$. Again a few calculations:

$\Box$ E!($u$) at @
$\forall w^{@\mathrm{A}w}$(E!($u$) at $w$)
$\forall w^{@\mathrm{A}w}(\Diamond$Int($u$) at $w)$

yield:

[□ E*x*@] □ E!(*u*) at @ ↔ $\forall w^{@\mathrm{A}w}(\exists w_1{}^{w\mathrm{A}w_1}(\mathrm{Int}(u) \text{ at } w_1))$.

Superscripts '@A*w*' and '*w*A$w_1$' state, respectively, that *w* must be accessible from @ and that $w_1$ must be accessible from *w*. As I mentioned in the Introduction, in the next two sections (3 and 4) I will show that two accessibility relations are available in Husserlian idealism. In Section 5 I will argue that one relation gives necessarily existing universals, while the other gives contingently existing universals. In other words, depending on the chosen accessibility relation I will derive:

[Nec] □ E!(*u*) at @

in one case, and:

[Cont] ¬□ E!(*u*) at @

in the other.

## 3. World-Bound and Free Imagination

I have mentioned that part of my regimentation of Husserl's view that it makes perfect sense to speak of contingent universals consists in understanding accessibility—the A relation of Section 2—in terms of imagination. In this section I briefly present Husserl's two notions of imagination; in the next I discuss them qua accessibilities. The most instructive text to look at in this connection is Husserl 2005. Useful remarks may also be found in Husserl 2013.

As I suggested, and as Wallner 2014 argues at length, Husserl was an actualist of sorts: he thought that possible worlds, and possibilities in general, should not be understood as non-actual counterparts of our world existing just as concretely as it does. For Husserl, possibilities are ideal objects (Husserl 2001) and, in particular, they are the intentional objects of acts of imagination:

> A possibility is posited when anything at all with such and such a sense is posited as something that can be realised by phantasy intuition (Husserl 2005: 661).

The notion that the achievements of our mental performances should be a reliable way of tracking possibility has been doubted by several influential analytic philosophers (most famously Yablo 1993). Husserl's view must thus be anathema for those philosophers: because for Husserl not only does imagination track possibility, but—as he would put it—'constitutes' it. In the Husserlian framework, however, this is no problem at all: because in it everything is, at some level, an achievement of our mental performances.

Possibilities, and possible worlds, are thus the intentional objects of acts of imagination. Imagination itself comes for Husserl in two species: world-bound and free imagination (*freie Phantasie*). The former term, I should mention, is not Husserl's, because—to the best of my knowledge—he only refers to it as 'phantasy' without qualifying it and rather relying on the context. World-bound imagination yields real possibility (*reale Möglichkeit*), while free imagination yields pure possibility (*reine Möglichkeit*). I have no space to follow Husserl's extremely detailed and fascinating descriptions and characterisations. They are in Husserl 2005, especially No 19 and Appendix LVII. The upshot, however, is the following.

World-bound imagination is constrained by what is the case in the actual world, while free imagination is not. Suppose, for example, that upon looking at

a red house I imagine it suddenly turning blue (not my suddenly seeing it blue, but its suddenly turning blue). In world-bound imagination, the transformation is imagined but then 'cancelled' and deemed impossible: because the actual world speaks against the possibility of the house suddenly turning blue. As Husserl puts it, the red 'raises a protest against the house being blue', as the house's suddenly turning blue 'has no motivation in actual experience [and thus in the actual world]' (Husserl 2005: 640). The same happens if, upon seeing a pen on a table, we world-bound imagine picking it up, dropping it and that it should float towards the ceiling: the actual world speaks against that, and so the imagined event is deemed impossible—or, more precisely, really-impossible (i.e., not a real possibility). Again Husserl:

> What I have given as existing […] I can 'imagine as otherwise'. I can phantasy it as if it were otherwise. I can suppose, assume hypothetically, that it is otherwise. The supposition is then […] abrogated as null by my actual experience (Husserl 2005: 674).

Not so, however, in free imagination, which is such that the actual world has no hold on it. So I can imagine jumping from my balcony and freely fly over London, and have tea with Mary Poppins on a cloud. Of course the actual world speaks against the event; but in free imagination that is irrelevant. So what distinguishes the two types of imagination is whether, while engaging in them, we hold on to the belief of the world and its implications for possibility or suspend it. Another way of putting it is indeed (to coin a phrase) in terms of suspension of disbelief: a disbelief motivated by what the actual world looks like.

This difference in imagination carries over to possibility: what is really possible with respect to the actual world is constrained by what the actual world looks like, whereas what is purely possible is not. Again, features of imagination carry over to possibility because of the peculiar framework we are in: Husserlian intentionalism, idealism and phenomenological actualism. The move would be suspicious in a realist environment.

Now, accessibility is precisely designed to capture semantically the notion of possibility with respect to a world and make it precise. A possible world's being accessible from another is simply the former's world being possible with respect to the latter. It is just natural, then, to treat the two notions of imagination as two distinct accessibility relations in the Husserlian framework. Importantly, they are both legitimate: we can engage in each of them more or less at will, and, although their results are sometimes incompatible, the acts themselves are not.[6]

Finally, that precisely these notions of possibility are at play when, in Husserl's philosophy, it comes to universals, is readily substantiated. Here is a passage from *Experience and Judgement*:

> The extension of [empirical] concepts is indeed infinite, but it is an […] extension […] of things […] really possible in [I read 'in' as 'with respect to'] the given world. These real possibilities […] must not be confused with the pure possibilities to which pure generalities refer (Husserl 1973: 330).

[6] 'Sometimes incompatible', as opposed to 'always', because I can very well freely imagine something that the actual world does not speak against.

In the next section I will first discuss the structural properties of accessibility in an abstract fashion, and then bring in world-bound and free imagination and suggest which of those properties they enjoy.

## 4. Structural Properties of World-Bound and Free Imagination

Accessibility relations are liable to have a number of structural properties. If $w_1$, $w_2$, $w_3$ are possible worlds, then:

*Reflexivity*: $w_1 A w_1$
*Symmetry*: If $w_1 A w_2$, then $w_2 A w_1$
*Transitivity*: If $w_1 A w_2$ and $w_2 A w_3$, then $w_1 A w_3$
*Left-Euclideanity*: If $w_1 A w_3$ and $w_2 A w_3$, then $w_1 A w_2$
*Right-Euclideanity*: If $w_1 A w_2$ and $w_1 A w_3$, then $w_2 A w_3$
*Seriality*: For every $w_1$, there is a $w_2$ such that $w_1 A w_2$

Notice that Euclideanity (either right- or left-, or both) entails transitivity, and reflexivity and Euclideanity together entail transitivity and symmetry. In **S4**, a normal modal logic, accessibility is reflexive and transitive. In **S5**, the most powerful and—I believe—most widely endorsed modal logic, accessibility is reflexive, symmetric and transitive (or reflexive and Euclidean).

I suggest that world-bound imagination is both reflexive and serial, but neither symmetric nor Euclidean (neither left- nor right-Euclidean). I will argue as much in Section 5 as part of my proof of [Cont]. What this means structurally, however, is that one can always world-bound imagine a world $w$ from that same world $w$ (reflexivity), and that there is always a world $w_1$ that can be world-bound imagined from $w$ (seriality). However, and again at this stage this is only a suggestion, it is not always the case that if $w_1$ is world-bound imaginable from $w$ then $w$ is also world-bound imaginable from $w_1$ (non-symmetry). Moreover, it is not the case that if a world is world-bound imaginable from two worlds, then one of these two worlds is world-bound imaginable from the other (non-left-Euclideanity). Finally, it is not the case that if two worlds are world-bound imaginable from a third one, then one of the two is world-bound imaginable from the other (non-right-Euclideanity). Free imagination, on the other hand, has no constraints at all, and thus, I suggest, cannot be anything short of reflexive, symmetric and transitive (or reflexive and Euclidean). This seems to me to follow straightforwardly from the notion, and therefore I will assume it rather than arguing for it. It is crucial, however, for my proof of [Nec].[7]

Let us map this onto logic. If I am right and world-bound imagination is neither reflexive nor Euclidean, then it has no place in **S5**. As a corollary, real-possibility is not **S5**-like. On the other hand, since free imagination is reflexive, symmetric and transitive (or reflexive and Euclidean); it is an **S5**-accessibility relation; as a corollary, pure possibility is **S5**-like.

It is an interesting question whether world-bound imagination is at least an **S4**-compatible accessibility relation: for accessibility in **S4** need not be either symmetric or Euclidean. It only needs to be reflexive and transitive. My conjecture is

[7] Actually free imagination is, for Husserl, constrained by the essence of the imagined objects. This connects with Husserl's theory that essence is the source of necessity and to his method for discovering facts about essence, 'free imaginative variation'. It is, however, beyond my purpose here.

that world-bound imagination is not transitive, because even if nothing in $w_1$ speaks against $w_2$ and nothing in $w_2$ speaks against $w_3$—so that $w_2$ and $w_3$ are world-bound-imaginable from $w_1$ and $w_2$ respectively—there may still be something in $w_1$ that does speak against $w_3$: something which is in $w_1$ but not in $w_2$. The catch is that this something, present in $w_1$ and lacking in $w_2$, must be such that its absence from $w_2$ is no incompatible with $w_1$: because otherwise $w_2$ would not, after all, be world-bound-imaginable from $w_1$. I have to say I have not been able to construct a plausible situation; so as far as I am concerned the notion that world-bound imagination is not transitive is mere conjecture. In any event, it has no role to play in what follows.

## 5. Derivation of [Nec] and [Con]

Assume Int($u$) at @ (i.e., that universal $u$ is an intentional object in the actual world). If, as seems plausible, world-bound imagination is reflexive (a world is world-bound-imaginable from itself), there is at least a world $w$ accessible (world-bound imaginable) from @ in which Int($u$), namely, @ itself. Thus,

(1) ◊Int($u$) at @.

This is a rather trivial way of securing (1). Here is a more substantial way. Surely if Int($u$) at @ there is always a $w \neq @$ that is world-bound-imaginable from @ such that Int($u$) is true in it (this, among other things, shows that world-bound imagination is serial). Because if $u$ is an intentional object at @, then it has a genetic basis $a$, $b$, $c$ in @, which in turn means that there are $u$-particulars in @. But *this* means that, since actuality entails possibility, $u$-particulars are not only actual but possible and, in particular, really-possible and world-bound-imaginable from @. For if they were not, then the actual world would speak against their possibility—except that it does not, because, by hypothesis, there simply are $u$-particulars in the actual world. Thus, there is a $w$ such that $u$ is constituted in it on genetic basis $d$, $e$, $f$ (where $d$, $e$, $f$ may or may not be identical with $a$, $b$, $c$). But if Int($u$) in $w$, then, since $w$ is world-bound-imaginable from @, (1) will follow non-trivially.

Once we have (1),

(2) E!($u$) in @

is secured. But how does $u$ exist at @? Necessarily or contingently? Let us first prove:

[Cont] ¬E!($u$) at @.

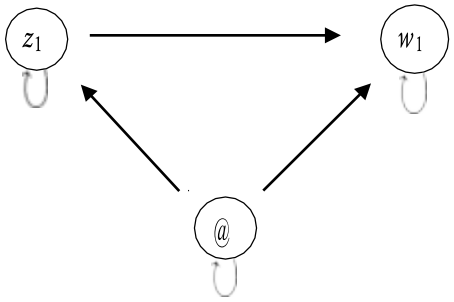
For [Cont] to be true, there must be a $w_1$ world-bound-imaginable from @ such that ¬E!($u$) at $w_1$; that is, a $w_1$ world-bound-imaginable from @ such that ¬◊Int($u$) at $w_1$. In turn, this means that there must be a $w_1$ world-bound-imaginable from @ such that no $w_2$ world-bound-imaginable from $w_1$ is such that Int($u$) in $w_2$. I will now try to construct $w_2$. Whilst doing so, I will in effect be giving a counterexample to the symmetry and the Euclideanity of world-bound imagination.

Consider a $w_1$ with no canids: no dogs, no wolves, no jackals and so on. This world is world-bound-imaginable from the actual world: because even if there actually are canids, it is not incompatible with this fact that there might not be. Now put yourself in the shoes of a citizen of $w_1$. Could you world-bound-imagine canids? I submit you could not: just as we cannot world-bound-imagine dragons, say, or chimerae, because our own world speaks against those kinds of animals.

Of course, just as in our world we have had a Hieronymus Bosch who imagined (and depicted!) all sorts of weird animals (e.g. in The Garden of Earthly Delights, especially the Hell panel), there could be a Bosch in $w_1$ able to imagine (and depict) canids. That, however, would be free, not world-bound imagination: otherwise, our Bosch's weird animals would be world-bound-imaginable from our world, and that is not the case.

If I am right, then, $w_1$ is such that no $w_2$ with canids in it is world-bound-imaginable. In a world where no canids dwell, however, the universal Canid has no genetic basis and therefore is not an intentional object. Thus, in no $w_2$ world-bound-imaginable from $w_1$ (including $w_1$ itself)) is $u$ an intentional object. Thus, $\neg\Diamond\text{Int}(u)$ at $w_1$. Hence, there is a $w_1$ world-bound-imaginable from @ in which $\neg\Diamond\text{Int}(u)$. As a consequence, there is a world world-bound-imaginable from @, namely $w_1$, in which $\neg\text{E!}(u)$. Hence, $\neg\Box\,\text{E!}(u)$ at @: [Cont] is proved and $u$—the universal Canid, in our example—is contingent in the actual world.

A diagram may help visualise how this situation is a counterexample to the symmetry and the Euclideanity of world-bound imagination. Here is a restriction to worlds @, $z_1$, $w_1$ of the graph of A (for A = word-bound imagination):




@ is the actual world, $w_1$ (from the proof above) is a world world-bound imaginable from @ with no canids in it, and $z_1$ is a world world-bound imaginable from @ with canids in it. The arrows represent both accessibility and the fact that accessibility is reflexive (loops) and non-symmetric. The latter, notice, is readily shown: we have $@\text{A}w_1$ but $\neg w_1\text{A}@$, because no world with canids is world-bound imaginable from a world with no canids. Against left-Euclideanity, notice that $w_1Aw_1$ and $z_1\text{A}w_1$, but, for the same reason, $\neg w_1\text{A}z_1$. Finally, against right-Euclideanity, notice that $@\text{A}w_1$ and $@\text{A}z_1$, but, again for the familiar reason, $\neg w_1\text{A}z_1$.[8]

If accessibility is free imagination, on the other hand, $\neg\Box\,\text{E!}(u)$ at @ cannot be derived: because $w_2$ cannot be constructed. The reason is that free imagination from $w_1$ does not depend on what $w_1$ looks like. Therefore, canids will be freely imaginable in $w_1$ even if no canids have ever existed in $w_1$. Every world is freely imaginable from any world. In terms of accessibility, this means—as I have already mentioned—that free imagination is reflexive, symmetrical and transitive, and pure possibility is **S5**-like. But then the **S5** axiom:

[8] Since $z_1$ has canids in it, worlds with canids, including @, are—barring other incompatibilities—accessible from it. This does not show in the diagram; on the other hand, it is irrelevant to my purpose.

(S5) $\Diamond p \rightarrow \Box \Diamond p$

will apply, guaranteeing that every possibility is a necessary possibility. Coupled with:

(1) $\Diamond$Int($u$) at @.

this gives [Nec] straightforwardly.

A final remark. I have claimed that even though $\Diamond$Int($u$) is a necessary but not sufficient existence condition for $u$, treating it as necessary and sufficient (as I have done) is not problematic for my argument. It is high time I said why. My strategy to prove [Cont] has been the following:

$\Box$ E!($u$) at @ $\leftrightarrow$ $\Box \Diamond$Int($u$) at @
$\neg\Box \Diamond$Int($u$) at @ (with WBI)
$\therefore$ $\neg\Box$ E!($u$) at @

In other words, I effectively worked on [$\Box$ Ex@] to negate $\Box$ E!($u$) at @. Now, treating $\Diamond$Int($u$) as a merely necessary condition just means formulating [$\Box$ E*x*@] as a conditional rather than as an equivalence. But this would still suit me, because to implement the strategy—working on [$\Box$ Ex@] to negate $\Box$ E!($u$) at @—I do not need an equivalence: a conditional will do. I used an equivalence rather than an implication only for the sake of simplicity: because that way I was able to use substitution instead of the somewhat more laborious modus tollens (or contraposition and modus ponens).[9]

## References

Drummond, J. 1990, *Husserlian Intentionality and Non-Foundational Realism: Noema and Object*, Dodrecht: Kluwer.

Fine, K. 2005a, "Plantinga on the Reduction of Possibilist Discourse", in *Modality and Tense*, Oxford: Oxford University Press), 176-213. Originally published in Tomberlin, J.E. and van Inwagen, P. (eds), *Alvin Plantinga*, Dordrecht: Reidel, 145-86.

Fine, K. 2005b, "The Problem of Possibilia", in *Modality and Tense*, Oxford: Oxford University Press, 214-34. Originally published in Loux, M. and Zimmerman, D. (eds), 2003, *The Oxford Handbook of Metaphysics*, Oxford: Oxford University Press, 161-79.

Hintikka, J. 1975, *The Intentions of Intentionality*, Dodrecht: Reidel.

Hintikka, J. and Harvey, C. 1992, "Modalizations and Modality", in Seebom, T. et al. (eds.), *Phenomenology and the Formal Sciences*, Dodrecht: Kluwer, 59-77.

Hopkins, B. 2016, "Numerical Identity and the Constitution of Transcendence in Transcendental Phenomenology", *Research in Phenomenology*, 46, 205-20.

Husserl, E. 1959, *Erste Philosophie (1923/1924). Zweiter Teil: Theorie der phänomenologische Reduktion*, The Hague: Martinus Nijhoff.

Husserl, E. 1969, *Formal and Transcendental Logic*, The Hague: Martinus Nijhoff.

Husserl, E. 1973, *Experience and Judgement*, Evanston: Northwestern University Press.

Husserl, E. 2001, *Logical Investigations (2 vols)*, London: Routledge.

[9] I am grateful to David Smith and Alberto Voltolini for many helpful comments. I am also grateful to Sara Gazo for proofreading and for helping me come to grips with imagination.

Husserl, E. 2005, *Phantasy, Image Consciousness, and Memory (1898-1925)*, Dordrecht: Springer.

Husserl, E. 2013, *Zur Lehre vom Wesen und zur Methode der eidetischen Variation. Texte aus dem Nachlass (1891-1935)*, Dordrecht: Springer.

Mohanty, J.N. 1985, "Intentionality and 'possible worlds'", in *The Possibility of Transcendental Philosophy*, Dordrecht: Martinus Nijhoff, 25-44.

Mohanty, J.N. 1999, "Phenomenology and the Modalities", in *Logic, Truth and the Modalities from a Phenomenological Perspective*, Dordrecht: Springer, 168-79.

Searle, J. 1983, *Intentionality. An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press.

Smith, A.D. 2002, *The Problem of Perception*, Cambridge (MA): Harvard University Press.

Smith, A.D. 2003, *Husserl and the Cartesian Meditations*, London: Routledge.

Smith, D.W. 2007, *Husserl*, New York: Routledge.

Smith, D.W. and McIntyre, R. 1982, *Husserl and Intentionality. A Study of Mind, Meaning and Language*, Dordrecht: Reidel.

Sowa, R. 2007, "Essences and Eidetic Laws in Edmund Husserl's Descriptive Eidetics", *New Yearbook for Phenomenology and Phenomenological Philosophy*, 7, 77-108.

Spinelli, N. 2016, "What It Is to Be an Intentional Object", *Disputatio*, 8 (42), 93-112.

Van Atten, M. 2007, *Brouwer Meets Husserl. On the Phenomenology of Choice Sequences*, Dordrecht: Springer.

Van Atten, M. 2015, *Essays on Gödel's Reception of Leibniz, Husserl, and Brouwer*, Dordrecht: Springer.

Wallner, M. 2014, "Phenomenological Actualism. A Husserlian Metaphysics of Modality?", in Rinofner-Kreidl, S. and Wiltsche, H.A. (eds.), *Analytic and Continental Philosophy: Methods and Perspectives*, Berlin: de Gruyter, 283-85.

Welton, D. 2000, *The Other Husserl*, Bloomington: Indiana University Press.

Yablo, S. 1993, "Is Conceivability a Guide to Possibility?", *Philosophy and Phenomenological Research*, 53 (1), 1-42.

# Intentional Relations

*Mark Sainsbury*

*University of Texas at Austin*

*Abstract*

Thinking about Obama and thinking about Pegasus seem to be the same kind of thing: both are cases of thinking about something. But they also seem to be different kinds of thing, in that one is relational and the other not. This paper aims to show a way out of the impasse by distinguishing varieties of relationality, concluding that what matters is the two-term relational nature of all intentional states, regardless of whether or not the representations they involve have referents.

*Keywords*: Intentionality, Relationality, Nonexistents, Meinong, Intensionality.

## 1. A Problem About Relationality

In some intentional states, the mind is related to the world. When I think about Obama, I stand in a relation to him: the state I am in could not exist unless he exists. In other intentional states, this straightforward relationality is absent, for two familiar reasons: (a) my intentional state is directed at something that does not exist and (b) my intentional state is nonspecific. When I think about Pegasus, reason (a) says that I cannot be in a relational state: I am in a state that obtains even though Pegasus does not exist. When John wants a sloop, reason (b) says that he may fail to be in a relational state: although plenty of sloops exist, John may not have fixed his desire on any one of them. We are faced with a paradox: some intentional states are relational and some are not. But all intentional states are the same kind of thing, and things of the same kind are either all relational or all non-relational.[1] The aim of this paper is to show a way out of this seeming impasse.

I will take for granted that intentional states involve relations to mental representations.[2] We can divide such *states* according to the kind of representation

[1] Compare Prior (1971: 130): "(a) X's thinking of Y constitutes a relation between X and Y when Y exists, but (b) not when Y doesn't; but (c) X's thinking of Y is the same sort of thing whether Y exists or not. Something plainly has to be given up here; what will it be?"

[2] Fodor says "I seem to have grown old writing books defending RTMs [representational theories of mind]" (Fodor 1998: 1). The book from which this quotation is drawn (*Concepts*) is a good introduction. Other notable proponents of representational theories of mind include Dretske 1995, Lycan 1996, Harman 1973, Field 1978, Sterelny 1990.

*Argumenta* 2, 2 (2017): 327-339
ISSN 2465-2334


DOI 10.23811/49.arg2017.sai

involved, for example distinguishing intentional states in which the subject is related to a truth-evaluable conceptual representation, which I call a thought, from intentional states in which a subject is related to a conceptual representation which is not truth-evaluable. Intentional states in the first category are commonly called "propositional attitudes". Those in the second category, if indeed there are any, are sometimes called "objectual attitudes". To avoid potentially misleading associations of this terminology, I shall refer to the first category of intentional states as those involving thoughts, and the second category as those not involving thoughts. Intentional states that do not involve thoughts typically involve concepts, and concepts are mental representations which, when suitably combined, make up thoughts. All intentional states involve concepts, but some may involve a concept or conceptual structure which is not evaluable for truth, and so is not a thought. If you are asked to think of a number, it might be that the number nine comes to mind, and on one view there is nothing more to your intentional state, so far as its representation is concerned, than your exercise of the concept NINE. This concept is not evaluable as true or false, so this is a potential example of an intentional state not involving a thought.

We can divide *attributions* of intentional states into two kinds, often called clausal and non-clausal. A paradigm of a clausal, or as I shall say sentential, attribution consists of an intensional verb like "believes", followed by an optional "that", followed by a complete sentence. A paradigm of a non-clausal, or as I will say nonsentential, attribution consists of an "intensional transitive verb", followed by a noun phrase, as in "Raoul is thinking about Obama".[3]

The following pairing may seem natural, but I will avoid commitment to it: sentential attributions are made true by intentional states involving thoughts, and non-sentential attributions are made true by intentional states not involving thoughts. The reason not to take this pairing for granted is, as hinted earlier, that there is room for doubt whether there are any intentional states not involving thoughts. The skepticism might be grounded in two ways. (i) Perhaps non-sentential attributions are reducible in every case to sentential attributions. Since sentential attributions require as truth-makers intentional states involving thoughts, all truth-making intentional states involve thoughts. So there are no intentional states not involving thoughts (or at least none that we can ascribe). (ii) Just as Frege said that only in the context of a sentence does a word have meaning, so one might think that only in the context of a thought can a concept be exercised. This means that in every intentional state involving the exercise of concepts, a thought is involved. This metaphysical claim does not involve commitment to the semantic reducibility envisaged in (i).

Although there may be cases in which nonsentential attributions are equivalent to sentential ones, I reject the full generality claimed by (i) (along with Montague 2009, Forbes 2017 and many others). Rejecting (i) does not exclude the possibility that (ii) is correct, and I leave that possibility open. This disrupts the tidy association between, on the one hand, thought-involving intentional states and sentential attributions and, on the other, non-thought-involving intentional states and nonsentential attributions. If (ii) is true, the truth-makers for nonsentential attributions are thought-involving intentional states.

[3] Here I take for granted, for terminological convenience, that an expression of the form "that p" is not a noun phrase.

Given the rejection of (i), the semantics for nonsentential attributions is not going to be the same as for sentential attributions, and the question asked in the opening paragraph stands: do nonsentential attributions attribute relational states or not? We incline to an affirmative answer when we consider Raoul's thinking about Obama and a negative one when we consider his thinking about Pegasus.[4] Yet intuitively all intentional states are of a single kind, as Prior stressed.

Much of the history of discussions of intentionality could be described as attempts to come to terms with these seemingly conflicting features. I start by considering two ways in which one might attempt to reveal all intentional states as relational. According to one, every intentional state involves a relation to a mental object: an idea, a sense datum, an intentional inexistent (Brentano 1874/2009: 68), or whatever; this is discussed in §2. According to the other, discussed in §3, we can achieve uniform relationality in another way, by allowing our ontology to include nonexistent objects.

## 2. Mental Objects

Intentional states do indeed have a kind of uniform relationality: according to representational theories of mind, they all involve a relation between a subject and a representation. This is a feature of the metaphysical nature of intentional states. But attitude attributions do not report these relations, even if the relations must obtain for the attributions to be true. When we report Raoul's state as his thinking about Obama, we do not refer to the concept OBAMA. We exercise the concept without referring to it, just as we may use the word "Obama" without referring to the word. Likewise, although Raoul himself may exercise the concept OBAMA, he is not thinking about that concept; rather, he is thinking about Obama, a very different kind of thing. Representations are not normally what our intentional states are *about* (your current intentional state is unusual), even if every intentional state involves the exercise of a representation.

Truth-makers for attitude attributions are relational in a way that the attributions themselves are not. This is not special to attitude attributions. We find the same structure in, for example, the attribution of weights. A truth-maker will involve the local gravitational forces, but these are not normally mentioned in an attribution. My cat weighs 10 lbs. This is made true by facts about how far she would depress a certain kind of balance in my local gravitational field. She would not depress it so far on the moon where the gravitational forces are different. But the attribution does not refer to gravitational forces. Likewise the attribution "Raoul is thinking about Obama" does not say that Raoul is exercising the concept OBAMA, even though his doing so is typically[5] what makes the attribution true. Because the term of the relation is not made explicit in the attribution of thinking (or of weight), I call the relevant relationality "covert".

Although this position is a straightforward consequence of a representational theory of mind, it risks being misunderstood, and it also does not provide a full

[4] The same questions arise, although somewhat less directly, for sentential attributions. "Raoul thinks that Obama is president" seems to ensure that Raoul is related to Obama, but we cannot make an analogous claim concerning "Raoul thinks that Pegasus flies". I will discuss the issue only for nonsentential attributions.

[5] But not always. In an attribution, we can sometimes use a concept not used by the subject in order to report the subject's state. The constraints on these permitted divergences are varied: see Sainsbury 2017.

resolution of our initial problem. It risks being mistaken for the incorrect view that intentional states are *about* representations;[6] and it fails to address the intuition that thinking about Obama is relational in a way that thinking about Pegasus is not (see §5).

Hume said that "the slightest philosophy" teaches us that "nothing can ever be present to the mind but an image or perception" (*Enquiry* 12.9). This means that trees and apples cannot be present to the mind. Given his Copy Principle, the view made it difficult for him to explain how we could so much as have an idea of external and continued existences (a problem he grappled with rather heroically in *Treatise* 1.4.2). This exemplifies a disastrous turn that a representationalist view may take: instead of saying that the intentional states are about what their representations are about, the fatal temptation for British Empiricist thinkers (and others) is to regard the intentional states as about the representations ("ideas") themselves.

The same turn is found in other contexts. It is fairly widely agreed that for any veridical sensory experience it is possible for there to be a non-veridical one that is indistinguishable to the subject. Hence a veridical and an indistinguishable non-veridical experience have something in common. A representationalist might express the commonality by saying that the cases involve two token representations of the same narrowest qualitative type. There would be no harm in calling these representations "sense data", except that "sense data theorists" wish to say that sense data are the *objects* of experience: are what is seen or smelled. But representations cannot be seen or smelled. They neither react appropriately with light nor emit odiferous molecules. For representationalists, all intentional states, including perceptual states, are relational, but the representations are not the "objects" of the states in the sense of what the states are about. Rather, the representations are what bring represented objects "before the mind". Analogously, we see by using our eyes, but we do not see our eyes. Using our eyes does not make our vision indirect.

Here are two recent examples of the erroneous view that, for representationalists, representations are the objects of intentional states. Prior describes a view he will reject in these terms: "in thinking of anything at all, we thereby put ourselves into a relation, not with *that thing*, but only with an idea or what-have-you which in favorable cases may 'represent' a real thing but in unfavorable cases does not" (Prior 1971: 127). If one omitted the phrase "not with *that thing*, but only", this would be a rough description of the view I defend. But the needless additions show that Prior is determined to turn it into the erroneous view, which he rightly goes on to reject.

Mark Richard (2001: 104) describes hunting and other intentional states like this: "When I hunt, expect, worship, fear, or loathe, I am focused on some representation of mine that is 'the object of my activity'". A representation is centrally present in the mind of a hunter, but the hunter's focus is not on the representation. Rather, through the representation, the subject focuses on what he is hunting, expecting, or worshipping. Hunting a lion involves representing a lion, but what

[6] Similarly for thought-involving intentional states and their sentential attributions: the states are not about the thoughts, but are about what the thoughts are about; and a sentential attribution does not refer to the thought in the subject's state and say that the subject is related to it.

is hunted is a lion. A lion, and not a representation, is "the object of my activity" if anything is.

## 3. Nonexistent Objects, and the Problem of Non-Objects

According to views commonly labeled Meinongian, there are nonexistent objects. It might seem that these would permit a uniform account of intentional states: the states are always relational, though in some cases the second relatum is nonexistent.

Meinongian views have been criticized for ontological extravagance, for creating "ontological slums", for flying in the face of a sober sense of reality, and so on. It is hard to take these criticisms very seriously, unless they are coupled with something more substantive. If Meinongianism can provide a smooth and consistent account of intentionality, should we not regard that as a significant virtue, to be weighed against these alleged metaphysical "costs"?[7]

There is a less familiar problem. Some nonsentential attributions have complements that seem not to refer to objects of any sort, existent or nonexistent, for example:

1. Jane wants *more cookies.*
2. Tom simply wants *out.*
3. Felicity feared *something worse than mere words*.
4. Julie thought about *getting David to dance*.
5. Walter is worried about *what to do about the mortgage.*
6. Bill wondered *how to tie his bow tie*.
7. Harriet was afraid of *what Jules might do next*.
8. Lois liked *how Superman flew*.

The italicized phrases answer the question what the attributed states are about. In this sense, they specify the "objects" of the states. But it may seem hard to take these seriously as entities, existent or nonexistent; and so hard to take them seriously as objects, in the ordinary sense of this word. If we do not take them seriously as genuine objects, there is a wide range of nonsentential attributions which cannot be shown to be relational by appealing to non-existent objects as relata.

The problem could be labeled the "problem of non-objects"; or, to rework a phrase from Meinong (1904: 83), the problem of objects ("intentional objects") that are not objects (in the ordinary sense, i.e. entities). The problem could be overcome in one of two ways: (i) by showing that the italicized phrases do refer to entities (possibly nonexistent), though perhaps of unfamiliar ontological categories; (ii) by showing that the attributions are really sentential at the level of logical form, and so create no more of a problem than do sentential attributions in general. The approaches are consistent, and one may be appropriate for some cases, the other for others. I will use "Meinongianism" for the view that nonsentential attributions are uniformly relational. (This is historically a bit misleading:

[7] Meinongianism must of course resolve its internal problems. Russell accused Meinong of being committed, incoherently, to the existence of the existent round square. Meinong attributes his own response to his student Mally. See Jacquette 1996: 16. It involves a distinction between "nuclear" and "extranuclear" properties, subsequently refined by, for example, Parsons. See Parsons 1978. The distinction has prompted disagreements. It is not needed in Priest's version of Meinongianism, though Priest's commitment to open and impossible worlds and to paraconsistent logic may be offputting to some. See Priest 2005.

a true follower of Meinong should keep to strategy (i), the objects to include "objectives" or propositions.) It will recognize nonexistent objects among the relata, and also whatever objects need to be found in order to address the problem of non-objects.

Approach (ii) seems appropriate for examples (1) and (2). Wanting more cookies might be held to amount to wanting to have more cookies, so that (1) "really" has a fully sentential complement: Jane wants it to be the case that *she has more cookies*. Likewise (2) might be taken to be equivalent to: Tom simply wants it to be the case that *he is out*.

Approach (i) seems appropriate for examples (3) and (4). "Something worse than mere words" might be held to denote a type of event, a potential object of fear.[8] (4) is arguably equivalent to: Julie thought about the action-type *getting David to dance*.

The complements in (5) and (6) are naturally understood as introducing indirect questions. While some approaches to questions do not readily extend to indirect questions in intensional contexts (see Karttunen 1977), a quick fix is tempting. A question was churning in Walter's mind: what shall I do about the mortgage? So it is tempting to regard (5) as equivalent to:

9. Walter worried about *the question* what to do about the mortgage.

The existence of the relevant question is ensured by the final phrase of the attribution, which actually presents it (in grammatically indirect form). Applying the idea to (6) requires a slight modification, since questions cannot be wondered:

10. Bill wondered about *the question* how to tie his bow tie.

Questions, it may be claimed, are genuine entities, and are the referents of the noun phrase complements in (9) and (10). Any meaningful interrogative sentence expresses a question, so we could model questions as classes of semantically equivalent interrogative sentences. Questions come into existence when they are asked or considered; perhaps once they exist they are eternal. These metaphysics do not seem any weirder than the metaphysics of contracts or surveys. So this type-(i) approach seems promising with respect to (5) and (6).

Not all "wh"-clauses introduce indirect questions, and the envisaged proposal does not apply to those that do not. Superficially similar relative clauses require a different treatment, as in (7) and (8) above, repeated here:

7. Harriet was afraid of what Jules might do next.
8. Lois liked how Superman flew.

Harriet was not afraid of a question, and Lois did not like a question, but that should not be surprising: (7) and (8) use the "what" and the "how" to form genuine noun phrases, rather than indirect questions. "What Jules might do next" refers to the action-type, *Jules's next action*. "How Superman flew" refers to *a way of flying*, a property of Superman's flying. So if we are fairly relaxed about what is to count as an object, happy to include properties, questions, action-types, event-types and so on, it looks as if apparent non-object complements can be shown either to involve reference to an object, or else to be in reality sentential complements, just as the Meinongian hoped.

[8] More exactly, the best candidate would be a type of event-types, those that are worse than types consisting of mere words.

Although the examples make the Meinongian strategy seem promising, we should be cautious. Consider Sean, who is worried about how his performance will go. We can express his worry in terms of a question: he is asking himself "How will my performance go?". It does not follow that he is worried about a question. We can imagine him responding: "It's not the question that worries me, it's my performance". But we cannot interpret his state as simply a worry about his performance. He might have a worry about his performance by being worried whether he will be permitted to perform, while being confident about how his performance will go if he is permitted. The "how" must be part of the story. We cannot model this "how" in terms of properties of the performance, for it may be that no such property is before Sean's mind. Part of his worry is precisely that he does not know what these properties will be, that is, he does not know how he will perform.

The issue is analogous to one that affects the view that all sentential complements refer to propositions. A standard problem is that fearing that the world will end is very different from fearing the proposition that the world will end (a fear that, as Graeme Forbes nicely puts it, affects only the "unduly timorous"). Likewise, worrying about how his performance will go is different from worrying either about the performance or about the question how it will go.

With this in mind, we might revert to Walter's worry about the mortgage. He might object: "it's not a *question* I'm worried about—it's my *mortgage*". We can hear words like "question" as referring to interrogative sentences or speech acts. Thus understood, questions are ordinary objects, but on this understanding Walter is right to say that it is not a question he is worried about. His worry is about *what to do*. It may take the form of Walter asking himself the question "What shall I do?", but this question, considered as a sentence or speech act, is not what worries him. Although we can loosely use the word "question" to introduce what Walter was worried about, reflection suggests that, to the extent that a question is straightforwardly an object, a question is not really what he is worried about. Something similar applies to Bill's bow tie. In the most ordinary sense of "question", it is not a question that Bill wonders about, it is how to tie his bow tie. We have not entirely disposed of the problem of non-objects.

This is not a knock-down argument that the Meinongian cannot deal with the problem of non-objects, but it injects a note of caution. Moreover, and setting aside generalized hostility to nonexistents, Meinongianism is vulnerable to two methodological issues. One concerns the notion of logical form invoked in the transformation of a nonsentential attribution into a sentential one. The other concerns the considerable tolerance displayed in the notion of an object.

Two sentences may be logically equivalent yet differ in their logical form (e.g. "p" and "p or p"). If "being the logical form of" is to be a relation of use to semantics, a sentence's logical form needs to reveal its semantic mechanisms. It is not easy to say what further constraints this imposes beyond identity of truth conditions. If semantics are to reflect the psychological reality of interpretation, then the theorist needs to state some general principles for transforming an arbitrary sentence into its logical form, principles which give insight into the processing procedures interpreters actually adopt.[9] The upshot in the present context is that logical form is not lightly to be invoked, and any use made of this notion

[9] DRT theorists are explicit that what I have called general principles should be algorithms. See Kamp et al 2011.

needs to be backed up with some indication of the relevant general principles. The Meinongian's attempt to treat some seemingly nonsentential attributions as "really" sentential is much more demanding than merely finding truth-conditional equivalents.

The other general issue is the capacious notion of an object required by Meinongianism. If one makes no distinction between "thing" and "object", the initial puzzle will lack bite. Raoul thinks about *something* whether he thinks about Obama or about Pegasus, and worries about *something* whether he worries about the mortgage or about Pegasus. If somethings are things, then every case involves a relation to a thing, and nonsentential attributions are readily seen to belong to a common "something"-attributing kind. But this is unsatisfying: there still seems to be a significant difference between the Pegasus case and the Obama case, which we cannot readily express if we make "object" as embracing as "thing". We need to find a criterion of objecthood that will ensure that not all things are objects. This raises another challenge, to be addressed by substantive metaphysical considerations.

The considerable differences among the objects invoked by Meinongianism raise other concerns. Thinking about Obama is the same in one way as thinking about Pegasus, for, in Meinongian perspective, both are thoughts about objects. But it is different in another way, since one is a thought about an existent and the other a nonexistent object. Even if both are genuine objects, the difference is considerable: thinking about an existent object might be based on causal relations, but nonexistent objects are causally inert, so what underlies thinking about one of these must be radically different. Likewise, thinking about a question seems pretty different from thinking about a person, even if a suitable notion of *question* can be excogitated. The global uniformity may obscure differences that need to be recognized.

Finally, Meinongian theories have nothing to say about the non-relationality induced by nonspecificity, as Priest has pointed out (Priest 2005: 64). Even if we could explain a desire for eternal life in terms of a relation to a nonexistent, namely eternal life, that explanation does not work for a nonspecific desire for a sloop. It is not that what the subject desires is nonexistent (there exist plenty of sloops), but simply that there is no specific object he desires. One might attempt to invoke another unfamiliar object, the "existentially generic sloop" (as in Lewis 1970/1983: 218). But it is clear that one who nonspecifically desires a sloop does not desire any such object. Nonspecificity means that Meinongianism as such will not deliver the uniform account of intentionality we envisaged. True, it counts *more* nonsentential attributions as attributing relational states than do many other accounts, but it does not supply the resources to justify claiming that *all* nonsentential attributions are relational.

These considerations suggest that it is worth exploring alternatives to the Meinongian position concerning the apparent non-uniformity of relationality in nonsentential attributions. I shall consider different kinds of relationality: factual, semantic, phenomenal and metaphysical.

## 4. Factual Relationality

A fact is *n*-place relational just if it involves *n* terms. The word "relational" is often used for facts involving 2 or more terms, and one-term facts are simply called nonrelational.

The fact that John lives in Texas involves more than one term: John and Texas. The fact that Harry lives in London and Berlin involves three terms, Harry, London and Berlin. The fact that the twins live in Bombay and Calcutta involves four terms, the twins (2), Bombay, and Calcutta. We have facts about living involving different numbers of terms: 2, 3 and 4. As I shall say, we have different degrees of factual relationality.

These different degrees intuitively do not show that the facts are of fundamentally different kinds in the three examples. Intuitively, living is the same kind of thing whether you do it in two places or one, or whether or not another person does it too. We can apply the moral to thinking. Raoul's thinking about Obama and his thinking about Pegasus differ in their factual relationality, the former being factually 2-term and the latter factually 1-term, and no doubt this is what leads us to suppose there is some important difference in the two cases. But what the analogy with living shows is that differences of factual relationality do not run deep: they do not undermine our view that all cases of living are of fundamentally the same kind. Likewise, differences of factual relationality among cases of thinking should not undermine our view that they are all of fundamentally the same kind. Just as living is living, and so fundamentally the same kind of thing whether done in two places or one, so thinking is thinking in both the Obama case and the Pegasus case.

## 5. Semantic Relationality

A fact is *n*-place semantically relational iff it can be stated by a sentence dominated by an *n*-place verb, one that takes *n* noun phrases to make a sentence. "Runs" (nontransitive) is a 1-place verb: it takes one noun phrase to form a sentence, "Rupert runs". "Kisses" is 2-place. A hypothesis is that semantic relationality can be more significant than factual relationality. What makes living "the same sort of thing" whether you do it in one place or two is that the semantic degree of the verb "lives" is constant.

"Lives" is a one-place verb. There is no so-called phrasal verb "lives-in".[10] This is shown by the fact that it needs no "in" clause at all, as in:

11. John lives with his wife/within his means/for love/to ride.

"John lives in Texas" consists of the sentence "John lives" plus the adjunct "in Texas", a "prepositional phrase". The difference in factual relationality in the examples concerning John and Harry comes from what is in the adjuncts. "Lives" itself is one-place.[11]

Perhaps semantic relationality can provide a more significant species of uniformity, one which trumps the non-uniformity of factual relationality. "Thinks" is a one-place verb, shown by the fact that it needs no "of" or "about" adjunct at all. A conspicuous example is Descartes' cogito, and here are some others:

[10] Compare the disparaging remarks about whether there are phrasal verbs in Huddlestone and Pullum 2002: 274. However, "give in" is a good candidate for being a phrasal verb, as in "He was so insistent that eventually I had to give in". This is obviously very different from "lives in" as used in "John lives in Texas".

[11] The point depends only on the constancy of the degree of "lives", and so can be accepted even by those who believe that "lives" is 2-place.

12. She thinks in the bath/clearly and dispassionately/while running/positively/too much for her own good.

If there is an "about" phrase, there is no restriction on how many terms it can introduce (from zero on up). As far as semantic relationality goes, thinking about Pegasus, about Obama, or about Russell and Whitehead are all of the same kind. The cases differ in their factual relationality, being respectively 1-term, 2-term and 3-term, but these differences are adventitious relative to the nature of thought. This uniformity in semantic relationality between thinking about Obama and thinking about Pegasus arguably justifies treating both as of the same kind.

Although semantic relationality is a genuine phenomenon, it is not one that can play the kind of taxonomic role just envisaged. "John lives in Texas" is equivalent to "John inhabits Texas", but the verbs are of different degrees, "inhabits" being at least two-place. This delivers a difference in semantic relationality, but clearly this is of no significance to the nature of living or inhabiting. Similarly, insisting on the one-place character of "thinks" makes it different from the genuinely two-place ("transitive") intensional verbs like "admires", "fears" and "wants". Yet any discussion of intentional states and intensional language needs a taxonomy that groups together thinking, admiring, fearing, and wanting, at some significant level of generality, even if the verbs differ in their semantic degree.

## 6. Phenomenal Relationality

An intentional state is phenomenally $n$+1-term relational iff in being in the state, it is for the subject as if there are $n$ things that are before her mind. This is an attempt to provide a notion of relationality "from the subject's point of view", in the tradition begun by Brentano. Even for one who knows that Pegasus does not exist, in thinking about Pegasus it is as if there were one thing before her mind, so her state is 2-term phenomenally relational. Thinking about Obama has the same degree of phenomenal relationality, thus securing the uniformity that our initial puzzle called for. If she thinks about a unicorn and a centaur, this counts as 3-term phenomenally relational, since for her it is as if there were two things before her mind. Phenomenal relationality is independent of belief. Even a subject who thought that there was in reality no such person as Obama counts as in a 2-term phenomenally relational state when she thinks about him.

The counting gets problematic in various cases, especially when plurals are involved. What degree of relationality is involved in thinking about unicorns? If we say "two" (the subject, and then unicorns counted as single object) it becomes plain that we are not really counting *objects* coming before the mind, but rather the representations exercised in the intentional state. The concept UNICORNS is just one concept, even if it supposedly represents more than one unicorn. If we are counting a plural there needs to be more than one. If we are counting concepts, then we get to 1, but concepts are not what are before the mind; rather, they are the enablers, not themselves objects of thought.

If we say that in thinking about unicorns the subject is in a state with an indefinite degree of relationality >2, we locate indefiniteness in the world, rather than just in language. The sentence "There are several dogs in the yard" is indefinite with respect to how many dogs are said to be in the yard. But the yard contains a definite number of dogs, from zero on up. The indefiniteness is confined to language. In trying to excogitate phenomenal relationality, we would have no

way of keeping the indefiniteness so confined. The only definite number of unicorns is zero, and if this is used in our counting, the degree of phenomenal relationality <2, which conflicts with the plural "unicorns".

If I want a sloop, it may well be that it is not for me as if there were a sloop before my mind. That would be the specific case, and my desire might be nonspecific. Again the question of the degree of phenomenal relationality has no satisfactory answer.

Brentano said that "Every mental phenomenon includes something as object within itself" (1874/2009: 88). The best interpretation of the remark is that intentional mental phenomena involve a relation to a mental representation. The notion of phenomenal relationality might be a failed attempt to describe this essential notion.

## 7. Metaphysical Relationality

Factual relationality does not seem metaphysically deep. At the relevant level of generality, *living* in Texas is the same kind of thing as *living* in London and Berlin. Likewise, *thinking* about Russell and Whitehead seems to be the same kind of thing as *thinking* about Obama. The difference between one and two "objects of thought" is not specially significant. Analogously, the difference between zero (the Pegasus case) and one (the Obama case) should not be regarded as striking.

The zero case is special in a different way, because it draws attention to the question how a nonsentential attribution can be true if there is nothing to which the noun phrase in the complement refers. This question does not arise when we consider the difference between one "object" and two, but it may arise when we consider the difference between zero "objects" and one, and this may explain why this seems like a special case: it dramatically reveals a core feature of intensionality. One explanation of this feature can be given by treating the complement position of intensional verbs as "semi-quotational": instead of the words being used in their normal committal way, they are put on display as a way of revealing features of the subject's intentional states. The truth of an attribution requires, not that the noun phrase in the complement refer to the right object, but that it express the right concept, one that reveals the nature of the subject's intentional state.[12] In a theory of this kind, there is nothing semantically problematic about the case in which the noun phrase fails to refer. The complements contribute to truth conditions in just the same way whether they refer or not.

Semantic relationality varies between semantically equivalent sentences (e.g. one constructed from the one-place "lives" and one constructed from the two-place "inhabits") and also varies between intensional verbs ("thinks" is one-place, "fears", "wants" and "admires" are two-place), whereas if our topic is intentionality we should keep these verbs in a single category. Phenomenal relationality seems unable to provide good answers about the degree of relationality in many cases.

[12] See the "display theory" first sketched by Sainsbury 2012, and Sainsbury and Tye 2013. A more developed version is by Sainsbury 2017. The idea goes back at least to Buridan (SDD 4.3.8-4): "talia verba [viz. intensional verbs] faciunt terminos sequentes appellare suas rationes" [make the terms that follow them invoke their meanings].

The one solid foundation is the two-place metaphysical relationality that is involved in all intentional states: a relation between a subject and a representation. Two things have led to confusion. One is the covert nature of this metaphysical fact. We do not *state* that this relation obtains when we make nonsentential attributions. Rather, the obtaining of some relation of that kind is what is needed to make the attribution true. The relationality is of a kind with the relationality of attributions of weight with respect to the local gravitational field.

The other source of confusion is, as already mentioned, between the representation and what the representation is about. Intentional states are not normally about the representations they exercise. The representation is not the state's "object", as that is often used. Rather, the state's object is whatever, if anything, the representation refers to, or is about. The notion of "aboutness" needed to make this true is itself intensional: a representation may be about Pegasus, and a thought about Pegasus involves a representation about him.

Metaphysical relationality is the fundamental feature of intentional states, the nature they all share. In the original puzzle, it was claimed that Raoul's thinking about Pegasus is not relational, since there is no such thing as Pegasus, whereas his thinking about Obama is relational, since there is such a thing as Obama. But in both cases the claims are made true by Raoul being in a two-place relational state, involving a Pegasus-representation in one case and an Obama-representation in the other. The metaphysical underpinning of thinking about Pegasus is just as relational as his thinking about Obama. For the Pegasus case, that is not because there really is such a nonexistent object as Pegasus, but because the truth-making state is a relational one, holding between Raoul and, in the typical case, the concept PEGASUS. For the Obama case, the state is relational in the relevant way not because there is such an object as Obama, but because the truth-making state is a relational one, holding between Raoul and, in the typical case, the concept OBAMA.[13]

## References

Brentano, F. 1874, *Psychology from an Empirical Standpoint*, McAlister, L.L. (ed.), London and New York: Routledge, 1973/2009.

Buridan, J. (SDD), *Summa de Dialectica*. Tr. G. Klima. New Haven and London: Yale University Press, 2001.

Davidson, D. 1967, "The Logical Form of Action Sentences", in N. Rescher (ed.), *The Logic of Decision and Action*, repr. in *Essays on Actions and Events*, Oxford: Oxford University Press, 1980, 105-21.

Dretske, F. 1995, *Naturalizing the Mind*, Cambridge (MA): The MIT Press.

Field, H. 1978, "Mental Representation", *Erkenntnis*, 13, 9-61.

Fodor, J. 1998, *Concepts*, Oxford: Clarendon Press.

Forbes, G. 2017, "Content and Theme in Attitude Ascriptions", in Grzankowski, A. and Montague, M. (eds.), *Non-Propositional Intentionality*, Oxford: Oxford University Press.

Harman, G. 1973. *Thought*, Princeton: Princeton University Press.

[13] Many thanks to Alberto Voltolini for comments on an earlier draft.

Huddleston, R. and Pullum, G.K. 2002, *The Cambridge Grammar of the English Language*, Cambridge: Cambridge University Press.

Hume. D. 1739-40, *A Treatise of Human Nature*.

Hume, D. 1748, *Enquiry Concerning Human Understanding*.

Jacquette, D. 1996, *Meinongian Logic. The Semantics of Existence and Nonexistence*, Berlin: de Gruyter.

Kamp, H., van Genabith, J. & Reyle, U. 2011, "Discourse Representation Theory. An Updated Survey", in Gabbay, D. (ed.), *Handbook of Philosophical Logic*, 2nd ed., Vol. XV, 125-394.

Karttunen, L. 1977, "Syntax and Semantics of Questions", *Linguistics and Philosophy*, 1, 3-44.

Lewis, D. 1970, "General semantics", *Synthese*, 22, 18-67; repr. in *Philosophical Papers, I*, Oxford: Oxford University Press, 1983, 189-232.

Lycan, W.G. 1996, *Consciousness and Experience*, Cambridge (MA): The MIT Press.

Meinong, A. 1904, "Über Gegenstandstheorie", in Meinong, A. (ed.), *Untersuchungen zur Gegenstandtheorie und Psychologie*, Leipzig: Barth, 1-50; Transl. "The Theory of Objects", in Chisholm, R.M. (ed.), *Realism and the Background of Phenomenology*, Glencoe (IL): Free Press, 1960, 76-117.

Montague, M. 2009, "Against Propositionalism", *Noûs*, 41, 503-18.

Parsons, T. 1978, "Nuclear and Extranuclear Properties, Meinong and Leibniz", *Noûs*, 12, 137-51.

Priest, G. 2005, *Towards Non-being. The Logic and Metaphysics of Intentionality*, Oxford: Clarendon Press.

Prior, A. 1971, *Objects of Thought*, Oxford: Clarendon Press.

Richard, M. 2001, "Seeking a Centaur, Adoring Adonis: Intensional Transitives and Empty Terms", *Midwest Studies in Philosophy*, 25, 103-27.

Sainsbury, M. 2012, "Representing Unicorns: How to Think About Intensionality", in Currie, G., Kotatko, P. & Pokorny, M. (eds.), *Mimesis: Metaphysics, Cognition, Pragmatics*, Vol. 17, London: College Publications, 106-31.

Sainsbury, M. 2017, "Attitudes on Display", in *Non-Propositional Intentionality*, Grzankowski, A. and Montague, M. (eds.), Oxford: Oxford University Press.

Sainsbury, M. and Tye, M. 2013, *Seven Puzzles of Thought and How to Solve Them: An Originalist Theory of Concepts*, Oxford: Oxford University Press.

Sterelny, K. 1990, *The Representational Theory of Mind: An Introduction*, Oxford: Wiley-Blackwell.

# Advisory Board

# The Tracking Dogma in the Philosophy of Emotion

*Talia Morag*

*Deakin University*

*Abstract*

Modern philosophy of emotion has been largely dominated by what I call the Tracking Dogma, according to which emotions aim at tracking "core relational themes," features of the environment that bear on our well-being (e.g. fear tracks dangers, anger tracks wrongs). The paper inquires into the empirical credentials of Strong and Weak versions of this dogma. I argue that there is currently insufficient scientific evidence in favor of the Tracking Dogma; and I show that there is a considerable weight of common knowledge against it. I conclude that most emotions are insensitive to the circumstances that might be thought to elicit them and often unfitting to the circumstances in which they arise. Taking Darwin's lessons seriously, even predictable emotional responses to biologically basic objects (e.g. bears, heights), should not be understood as tracking abstract categories (e.g. danger). This renders most contemporary accounts of emotion implausible. We are left with two options: one may still continue to claim that emotions aim at tracking, even if they often fail; or one may abandon the Tracking Dogma in favor of a non-representational view.

*Keywords*: emotions, Darwin, core relational themes, non-representational, empiricism

Modern philosophy of emotion has been conditioned in large part by a dogma, what I shall call the *Tracking Dogma*, according to which emotional reactions track features of the natural and social environment that relate to or bear on certain typical aspects of our well-being. As Annette Baier put it:

> *We all accept* the idea that emotions are reactions to matters of apparent importance to us: fear to danger, surprise to the unexpected, outrage to the insult, disgust to what will make us sick, envy to the more favored, gratitude for the benefactors […]. Emotion then plays the role of alerting us to something important to us—a danger, or an insult (Baier 2004: 200. *Emphasis* added).

In other words, every emotion-type such as fear or anger functions to track what a leading psychologist has called "core relational themes" (Lazarus 1991: 22) that match emotion types with roughly described types of circumstances, such

*Argumenta* 2, 2 (2017): 343-363
ISSN 2465-2334


DOI 10.23811/410.arg2017.mor

as fear with danger or threat, anger with a wrong we suffered, guilt with wrongs that we inflict, joy with benefit, pride with achievement, sadness with loss, jealousy with loss of affection. I use the term 'core relational theme' but these general descriptions of types of circumstances that matter to us are also known in the philosophical literature as the 'formal objects' of emotions.

Philosophers differ in how they conceive of core relational themes. Some think of them as conceptually structured aspects of the situation (e.g. *seeing* the situation *as* dangerous, thereby noticing its dangerous aspect).[1] Others (e.g. Brady 2007, Tappolet 2012) regard these themes as designating values (e.g. 'dangerous'), and yet others (e.g. Griffiths 1997) as features of the natural and social environment (e.g. danger, threat) that call for certain typical coping strategies, known as the emotion's action-tendencies, such as running away in fear or lashing out in anger. But they all presuppose that tracking is *objective* in the scientific sense, namely that an observer would identify certain circumstances as objectively dangerous to a person or animal with objectively appreciable needs and wants. Whether or not such objective dangers and wrongs etc. may be further reduced to some other kind of entities is a metaphysical question for those philosophers but whose answer is not relevant for the purposes of this paper.

I will use the term 'core relational themes' as well as 'tracking' without taking these differences of interpretation into account. Nor do I take into account the difference between speaking about general themes such as 'danger' or 'wrong' and speaking about how those themes can be broken down to components such as "goal relevance" or "coping potential" (Lazarus 1991: 39).[2] Such components or aspects also thematically characterize the way the environment bears on the subject's well-being. "The point is—as Jesse Prinz says—that core relational themes are directly relevant to our needs and interests" (Prinz 2004: 66). This is the crux of the dogma, shared by everyone who holds it, namely that the themes in question are *purposive*, that they directly bear on our well-being. I shall use the terms 'core relational themes' and 'purposive themes' interchangeably in the rest of this paper.

Core relational themes are matched to emotion-types creating "couples" such as [fear/danger] and [anger/wrong] or generally [E/T], where E stands for Emotion, T for Theme (and traditionally it has been said that T is the "formal object" of E). And these roughly described "couples" are the ones in reference to which we may judge whether or not the emotion is justified, that is, whether or not a given emotional reaction *fits* the situation in which it arose. Fear is a fitting response to dangerous situations; anger fits situations where we have been wronged, etc. These "couples" thus also articulate what D'Arms and Jacobson (2003: 132) call "norms of fittingness". According to the Tracking Dogma then, emotions *aim* at fittingness.

A tracking system can be characterized in reference to two empirical notions:

*Accuracy*: If it were not the case that the situation presents core relational theme T then the subject would not experience emotion E, where T and E belong to the above described 'couple' [E/T].

[1] These include 'quasi-judgmentalists' or the 'seeing-as' accounts such as Greenspan 1988. Some seeing-as account elaborate more on the conceptual structure of the perceived situation, e.g. Ronald de Sousa's (1987) "paradigm scenarios".

[2] For a summary of Lazarus's appraisal theory and the dimensions it involves, see Prinz (2004: 14-17).

*Sensitivity*: if the situation presents the purposive theme T then the subject experiences emotion E under normal conditions, where T and E belong to [E/T].

In other words, a sensitive tracking system would alert the subject whenever she faces dangers, wrongs, and other core relational themes in the nearby environment, and an accurate system will by and large "get it right." Whoever holds the Tracking Dogma is thus committed to one of its following versions: 1) The Strong Tracking Dogma: A *reliable* tracking system that is both sensitive and accurate. Such a system will allow for predictions as to when a subject will emotionally react and how. 2) The Weak Tracking Dogma: An emotional system that need not be sensitive to the instantiation of core relational themes but when it tracks it does so accurately. 3) The Normative Tracking Dogma: A system that *aims* at tracking core relational themes, but need not be either sensitive or accurate.[3]

In order to distinguish clearly between the Strong and the Weak Tracking Dogma, we need to know what 'normal conditions' amount to for the case of an emotional detection system. Indeed, all tracking accounts owe us a specification of what emotional normal conditions are in order to clarify how the tracking system is supposed to 'track'. If emotions do track core relational themes, then we have to assume there are such specifiable conditions and that the distinction between Strong, Weak and Normative versions of the tracking view holds.

That anyone who thinks that emotions *aim* at tracking core relational themes holds one of these views is a conceptual argument, which introduces a straightforward way to classify the many philosophical accounts of emotion and which further forces philosophers of emotion to clarify their commitments about the tracking systems they propose. But since each of these versions is an empirical hypothesis about the actual sensitivity and accuracy of the supposed tracking system, these views can be defended or criticized by turning to empirical support. This paper aims to cast serious doubt on the empirical plausibility of both sensitivity and accuracy. This doubt is therefore addressed at any view which takes either sensitivity or accuracy or both for granted, i.e. at the Strong and the Weak versions of the Tracking Dogma, which together comprise the majority of philosophers of emotion, whose prominent figures are identified in the first section. This leaves intact the possibility of the Normative version of the Tracking Dogma, which I do not criticize here. But I take myself as making more plausible an alternative to *any* version of the Tracking Dogma.

One main aim of this paper is methodological, namely to examine what kind of experience counts as evidence for or against the Tracking Dogma or any other general vision of what emotions are. In order to conduct such a methodological inquiry, we need a working definition for emotions that is as uncontroversial as possible, i.e. that is not theoretically biased. I propose this minimalist definition: *an episodic affective experience that is characterized by prototypical physiological and behavioral expressions, and that we often feel or experience as passively coming over us*.[4] And whatever the causal etiology of these affective episodes is, it differs

[3] I leave out the option of a sensitive and inaccurate tracking system, since it cannot be ascribed to any philosopher.

[4] Even when we attempt to induce a mood in ourselves by listening to music or watching the ocean, we do not actively order our emotion to appear but rather put ourselves in

from that of mere sensation (e.g. tummy ache caused by food poisoning). I do not assume that in having an emotion one needs to be aware that one does. That is, I allow that one may go through an emotional experience without awareness, e.g. be angry with someone while forcefully denying it and genuinely believing one is not angry. These affective episodes often involve attending to one's nearby environment, making people and things in the environment emotionally salient, namely more vivid and often targets of behaviors such as running away from them or attacking them. Anything further than this—such as claims about those emotionally salient objects being intentional objects or targets of tracking systems, or claims about emotions as providing us with information about the world that is beneficial in some way to our well-being—cannot be presupposed, but is rather a question about our experiences and practices.

So what kind of empirical knowledge about or epistemic access to our emotional experiences and practices can we avail ourselves to? This question becomes pertinent once we acknowledge that the scientific data often cited by philosophers of emotion cannot currently decide on the correct vision for emotions, as argued in the second section of this paper. The second section further argues that if we take Darwin seriously, then many co-variances between emotional responses and various objects that are not inherited from the evolution of our species (e.g. fear of exams, sadness caused by losing a job) should not be understood as tracking dangers or losses. Even scientifically verifiable co-variances between emotional responses and objects inherited from the evolution of our species (e.g. fear and bears), need not be understood as verification of tracking (e.g. danger). The third section argues that we may appeal to common knowledge that is based on our ordinary everyday experience of emotions, and the fourth section accordingly presents non-scientific empirical considerations against sensitivity and accuracy of the supposed tracking system.

I conclude with a brief introduction of a new vision for emotions, a rival to the only plausible option left from the Tracking Dogma, namely the Normative version of it. According to the new view, emotions do not aim at tracking dangers, wrongs, or any other purposive theme. Indeed, emotions do not have an intrinsic purpose that relates to our well-being. When emotions are triggered by objects such as bears or exams, they are not triggered by objects-qua-instantiating-a-purposive-theme such as danger. The triggering of emotions need not be law-like and the resulting emotions need not be fitting. I leave the criticism of the Normative Tracking Dogma and the full defense of the new alternative for another occasion.[5]

## 1. The Strong and the Weak Tracking Dogma: Definition and Supporters

According to the Strong Tracking Dogma, the emotional tracking system is presupposed to be reliable, namely both sensitive and accurate. The reliability claim is explicitly held by Jesse Prinz, one of the leading philosophers who regard

front of various familiar *triggers*, with the hope that they will work in the desired way again, a hope that is, by the way, not always fulfilled.

[5] For a critical discussion of appraisal theories of emotion, including those that qualify as the Normative Tracking Dogma, and a new view on emotions see Morag 2016.

emotions as produced by sub-personal law-like mechanisms.[6] Prinz articulates the Strong Tracking Dogma:

> Emotions are certainly set off by core relational themes. That is, they are reliably caused by relational properties that pertain to well-being (Prinz 2004: 66).

It is hard to find such an explicit declaration among the other sub-personalists. Yet their formulations indicate that they hold the Strong Dogma. Jenefer Robinson, for example, says that we are 'programmed' to emote in certain ways in the face of instantiated core relational themes. Unless the program is faulty, a 'program' assumes that a certain input will normally produce a certain output, that is, it assumes both sensitivity and accuracy.[7] Paul Griffiths talks about (basic) emotions as natural kinds that allow for "very reliable predictions" (Griffiths 2004a: 235).[8] In order to predict how one will emote in the face of what circumstances, the supposed tracking has to be both sensitive and accurate, at least most of the time. Everyone acknowledges that our emotions do not *always* fit the objects they make salient, but the Strong Tracking dogmatist considers such occasions to be divergences from normal functioning, "errors" (D'Arms 2000: 1468). Fallibility is in any case built-in to the empirical notion of reliability.

There is another camp in the philosophy of emotion that presupposes the reliability claim. Those are philosophers who regard emotions as a mode of perception.[9] Perceptual and perception-like capacities—which are representational and have correctness conditions—are presupposed to fulfill, by and large, their function. According to this model then, emotions "typically" (Brady 2007: 273) fit their targets, that is, they are by and large accurate. Indeed, accuracy is built into our language insofar as perceptual verbs like 'see', 'hear' and 'touch' have a success grammar. It makes no sense to say 'I see the tree in the garden' if there is no tree there to see. Furthermore, perceptual and perception-like states presuppose more than the possibility of being caused by the features they in turn represent. Under normal conditions (whatever those may be for the case of emotions), they are also supposed to be sensitive, to be reliably caused by those features of the environment that they aim to track. And there is also an ordinary expectation of sensitivity or perception, namely that if there is a tree in front of one under normal conditions that one will see it.

This is also compatible with the frequent analogy between recalcitrant emotions, those that do not dim down despite the subject's explicit judgment against them (e.g. phobias), and relatively rare 'optical illusions',[10] which suggests that

[6] Prominent sub-personalists are Prinz 2004; Griffiths 1997; Robinson 2005 and D'Arms & Jacobson 2003.

[7] Robinson allows for some biological pre-programming and further programming that depends on one's developmental history. See Robinson 2005: 63, 70, 71, 72, 74, 75.

[8] Griffiths speaks only of what he calls "basic emotions", which do not involve thoughts, but since the minimalist working definition of this paper does not include thoughts, this distinction is not relevant here.

[9] Philosophers who support a perceptual model for emotions include Charland 1995, Brady 2007, Tappolet 2012, Döring 2010, Debes 2009, Deonna 2006. Prinz is both a sub-personalist and an avowed perceptualist.

[10] Philosophers differ about how to unpack this analogy, but many of them refer to it, including (but not only) Strong Tracking Domgatists, such as Brady (2007), Prinz (2004: 235), Tappolet (2012) and Döring (2003: 223).

perceptualists regard recalcitrance as a marginal phenomenon, as appropriate for a perceptual model. Some philosophers who endorse the perceptual model hold at the same time that recalcitrant emotions "often" (Tappolet 2012: 210) occur, a claim which entails that we often emote unfittingly or that the supposed tracking system is inaccurate. To avoid confusion such philosophers should simply forgo the perceptual model and endorse the Normative Tracking Dogma that forgoes commitments to accuracy.

But to take the analogy with perception *seriously* is to accept the reliability claim. A perceptual system that often fails to detect what it is designed to detect (insensitive) or that often provides false information about the world (inaccurate) is a malfunctioning perceptual system. Presumably those who invoke the perceptual model to explain emotions are not to be credited with that. Emotions are thus understood by many philosophers of emotion to be a kind of a sixth sense: a "bodily radar" as it has been put (Prinz 2004: 240), reliably alerting and telling us where we stand in our natural and social environment.

Except for a few obvious cases such as fear of bears in the forest or anger when someone hits us, what counts as a wrong, as an achievement, as loss of affection, as a shameful failing or weakness or even as a danger and so on typically differs from one social niche to another. And social niches may be as small as we like. Supposing for the sake of argument that emotions track core relational themes then what counts as an instantiation of a core relational theme for a certain individual should by and large correspond to *that individual's endorsed normative standards of fit, qua a member of a certain social niche at a certain stage of life*. None of us invented those norms, but each of us may be said to hold slightly different standards of fit, depending on differences in how we understand or internalize them (cf. D'Arm & Jacobson 2003: 136).

This relativization of a tracking system to large and small social niches is surprisingly conceded by some perceptualists. Tappolet, an avowed perceptualist, calls it "plasticity" (Tappolet 2012: 220-21). Deonna calls it a perspectival "frame of reference" (Deonna 2006: 37).[11] Normal functioning perceptual capacities are nowhere near as variable as our emotional capacities.[12]

As for the sub-personalists, they typically accommodate the social relativity of the supposed reliable tracking system, by presupposing a developmental process of socialization, education and self-training that somehow feeds back into each individual's emotional system, which would then track the socially relevant dangers, wrongs, etc.[13] The attunement need not be one-directional whereby we change our emotional patterns in accordance to our normative judgments and education. In some cases we listen to our gut reactions rather than to our

[11] Deonna emphasizes that this frame of reference can be unique to an individual. However, since Deonna is committed to the tracking of values or "evaluative properties" (*ibid.*), which are at least in principle public, then for Deonna, this individual is a limit case of the smallest social niche whose other member is a possible, if not an actual other. Taking seriously the grammar of values means that at the very least, someone else at some point in time should be able to track the same values.

[12] For a comprehensive critique of the analogy between emotions and perceptions see Salmela 2011.

[13] See for example Prinz on re-calibration of "calibration files" which encode eliciting conditions that reliably cause emotions in Prinz (2004). And also D'Arms & Jacobson on "immunization" against un-fitting ("natural" or basic) emotions such as fear of harmless spiders, in D'Arms & Jacobson (2003: 144).

endorsed norms of fit, re-adjusting our norms to fit our emotions.[14] Although sub-personalists do not believe that each emotional reaction responds to reasoning, they typically assume some kind of interaction between one's normative and linguistic system and the emotional tracking system, an interaction that works in the long term, that adjusts emotional patterns over time. This is no small assumption, and phrases such as 'top down' do not comprise an explanation for this assumption, but I will not pursue this criticism here.

The basic idea is that, by and large, we are all brought up in light of the (albeit socially relative) Aristotelian ideal of the Phronimos, the well brought-up person who responds to situations with the fitting emotion-type.[15] This means that every *normal* human adult of a specific social niche would have law-like emotional patterns that reliably track what counts as dangers, wrongs, achievements, benefits and so on in her or his social niche.

Now the sub-personalist that allows for this top-down interaction, necessary to account for the plasticity of the wrongs, the dangers, the achievements etc. to which the supposed tracking system is sensitive, can take here two routes in understanding the term 'normal'. If 'normal' means 'normative', then it is possible for the educational and training developmental process to go well or not so well, and respectively it is possible that one's tracking system will be quite inaccurate. In this case, the sub-personalist may join the Normative Tracking camp. This, however, would force sub-personalists to re-think their view, since they would then have to say that their 'programs' to track dangers, wrongs etc. are faulty, and do not lend themselves to 'very reliable prediction'. The other option is to take 'normal' to mean 'statistically normal' and to maintain the socially-relative reliability claim for human adults.

The Weak Tracking Dogma holds that emotions may be insensitive to all kinds of wrongs, dangers, achievements etc., but that by and large and most of the time when we do emote the tracking system has succeeded in its aim to track purposive themes such as dangers and wrongs. That is, the Weak version forgoes sensitivity but still insists on accuracy. The main camp of philosophers of emotion that presupposes this view includes those who hold that emotions are modes of 'seeing-as', e.g. when one is afraid, one *sees* the situation *as* dangerous. The contemporary philosophers in the seeing-as camp are those who talk about purposive themes as involving concepts. That is, they claim one sees the situation in terms of the concept 'danger' or in terms of a conceptually structured danger-scenario.[16]

The seeing-as relation is famously demonstrated in Wittgenstein's example of the duck-rabbit drawing, an example often mentioned by these philosophers of emotion (e.g. Roberts 1988). When I see a duck in the duck-rabbit drawing, I

[14] This aspect of this developmental picture emerges from Justin D'Arms' discussion on empathy through contagion, in D'Arms 2000.

[15] Cf. Aristotle (1991: 1106b14-b21). In fact, the Phronimos is also said to emotionally respond at the right or fitting intensity. Interestingly, most tracking accounts either ignore or downplay the intensity factor (is it because of the prevalence of overreactions? Or is it because it is not clear whether overreactions matter and in what way to the evolution of our species?) It is possible to amend the above accuracy and sensitivity definitions to take account of intensity, but I shall not engage in this issue here, if only because of its relative absence from the current philosophical literature.

[16] Prominent seeing-as accounts include de Sousa 1987, Greenspan 1988, Rorty 1980, Lyons 1980 and Roberts 1988.

see the duck-aspect of the drawing, an aspect that is there to be seen. This seeing-as experience, as is commonly interpreted, requires me having the concept of a duck. Of course, I need not see the duck aspect. Similarly, according to seeing-as accounts, by and large when I see the situation as dangerous, the situation lends itself to the application of the concept 'danger' (accuracy), even if I need not see the situation in this manner in the first place (forgoing sensitivity). The accuracy claim is further implied by the fact that seeing-as philosophers regard recalcitrant emotions as "fringe cases" (Rorty 1980: 103).

## 2. The Lack of Scientific Evidence for Reliability of Tracking (Sensitivity and Accuracy)

Contemporary philosophers overwhelmingly use the term 'empirical support' to mean 'scientific evidence'. Indeed, sub-personalists often refer to scientific experiments to support their claims. In what follows, I summarize the experiments cited in the philosophical literature and then examine what they can be said to verify.

Some (disturbing) scientific evidence is presented for the reliability of certain typical newborn baby responses to certain typical circumstances.[17] Other experiments, such as those that show that we develop phobias to snakes much more easily than we develop phobias to flowers (Ohman, Fredrikson and Hugdahl 1976, cited in Griffiths 1997: 88), or that we very easily learn to fear spiders,[18] demonstrate perhaps that we all have a tendency to fear snakes and also spiders. Perhaps we also all have a tendency to be disgusted by cockroaches and fear earthworms (Griffiths 2004b: 95 and 1997: 28, 93). It seems plausible that newborns reliably emote in certain typical ways in response to the relatively limited set of types of objects and circumstances they relate to (e.g. the presence of a caregiver). But infants beyond the newborn stage already respond to more objects (e.g. they have favorite toys) and have past experiences that may shape or alter in some way or other their emotional responses. So whether or not or to what extent we can generalize from those experiments and claim that infants that are beyond the newborn stage are all more or less the same remains an open question.

What about human adults? One class of experiments makes use of emotionally laden memories. Some such experiments specifically ask people to recall life-changing events such as the death of a loved one, or to recall extreme emotional experiences.[19] If these experiments represent a sample of people's emotional life, then it is the one where people face what we may call 'significant' circumstances, such as a big failure, a grand success, one's wedding or

[17] Watson (1924: 229-31), Sroufe (1984: 112). More ethical experiments by Meltzoff and Moore 1977, Field *et al.* 1982, Meltzoff and Moore 1989 and 1995, Izard 1978, Trevarthen 1984. The first six experiments are cited in Robinson (2005: 37-38); the first three and the last two in Griffiths (1997: 88).

[18] Robinson (2005: 37) refers to Frijda's (1986) reports on many such studies.

[19] For example, Prinz (2004: 70-71) discusses an experiment where people who were injected with adrenaline were asked to recall life-changing events such as the death of a loved-one. He there cites Maranon 1924. See also Prinz's (2004: 96) discussion on experiments that ask people to recall occasions where they felt extreme guilt. He there cites Shin, Dougherty, Orr, Pitman, Lasko, Macklin, Alpert, Fischman & Rauch 2000.

graduation ceremony, the first time one saw the ocean, break-up conversations, historical events, etc.

Sub-personalists also tell us about Americans and Japanese that respond with disgust to gory films (Friesen 1972, cited in Robinson 2005: 34; Prinz 2004: 137) and about people getting stressed when subjected to films showing genital surgery (Lazarus & Alfert 1964, cited in Prinz 2004: 30). If we ignore gore fans, certain people with sadistic or masochistic tendencies and doctors who have become acclimatized to such things, then these experiments show that most of us reliably find the insides of bodies and the maiming of bodies disgusting and stressful. We also read about people being conditioned to dislike certain images via electric shocks.[20] Other experiments (Logue, Ophir, and Strauss 1986, cited in Griffiths 1997: 28) talk about disgust of foods that are associated with nausea via conditioning (even when one knows the nausea was not caused by the food). These experiments arguably show a predictability of response to circumstances associated with severe pain. These experiments may be said to be a representative sample of what we may call 'extreme' circumstances. The negative ones include events that we either prefer to avoid or are perversely fascinated with, e.g. rape, torture, a natural disaster, war, the big dipper at the Luna Park, bungee jumping, car crashes, open-heart surgery etc. The positive ones would be events such as winning the lottery.

Another class of experiments exposes subjects to basic facial expressions such as smiles, frowns and stares. Some experiments mentioned in the philosophical literature demonstrate, for example, that we reliably prefer images that we previously saw in conjunction with a smiley face (Murphy and Zajonc 1993, cited in Robinson 2005: 39-40). These experiments could be regarded as sampling the category we may call the 'clichéd'. The clichéd may also include the joy football fans feel when their team wins, and the warmth of endearment many people feel when they see babies or kittens.

I have not found in the philosophical literature experiments that test the disgust adults feel when exposed to vomit or rancid food, the startles we experience as a response to a sudden loud noise, the anger we feel when someone hits us, or other emotional responses to objects whose emotional import is clearly inherited from the evolution of our species. Call these the 'biologically basic' objects.

It seems plausible that the experiments cited in mainstream philosophy of emotion show that most of us reliably respond in certain typical and fitting ways to the extreme, the clichéd, and the biologically basic. There are reasons to doubt the experiments that regard significant events. People tend to report or think about themselves in a way that conforms to what they are expected to feel at their wedding, graduation ceremony, or their break-up, and people may recall events in the way that suits them, especially if they are practiced in recalling that specific event and describing and re-describing it. The responses to such events seem complex especially when we begin to describe them in the kind of detail that takes account of that person's biographical particularities. But it seems right that at some general level of description many of us feel sad when our loved one dies, joy and/or nervous when we get married, and have more or less predictable and fitting roughly-characterized emotional responses to events that clearly

[20] The experiment was conducted by R. Lazarus and R.A. McCleary in the 50's, described in Lazarus (1991: 155-56), cited by Robinson (2005: 40).

stand out in our life as more important than the rest of our everyday life. That most of us emote in predictable and fitting ways to the significant, the extreme, the clichéd and the biologically basic is also evident in the use of those kinds of circumstances in advertisement, soap operas, thrillers, the Luna park, and other commercial and 'formulaic' story-techniques that trade on the relative reliability of those kinds of circumstances.

Importantly, I do not claim that the abstract categories I articulated—the 'clichéd', the 'extreme', the 'significant' and 'the biologically basic'—are core relational themes. I am not claiming that they play any role in our psychology, that we track them under that description or any other, or that they qualify as an ecological category. Rather, I concede that there may be some causal co-variance to be found by science between typical emotional responses and some objects that can be judged to instantiate these abstract concepts, that some such identified co-variances can be further judged as fitting responses, and that people, as theorists with conceptual capacities, can thematize them, as I just did. The question now is: does this evidence of the co-variances described above count as evidence that emotions track purposive themes?

In what follows, I argue that even the emotional reactions to objects from these aforementioned categories do not track core relational themes within a theory of our well-being. Although the objects of the fear responses examined in those experiments can be described by an observer as dangerous, I suggest that those fear responses did not track *dangers*. Only when it comes to the 'biologically basic' objects, or the kind of responses we share with animals, does it make sense to claim that emotions track core relational themes, whose action-tendencies appear purposive, such as running away in fear from a bear in the forest. Even then, I argue, it is neither necessary nor explanatorily fruitful to hold the Tracking Dogma in any of its versions.

It is crucial to remember that the kind of purposiveness Darwin talked about with respect to evolved systems was a purposiveness *without a purpose*. Biological systems may appear *as if* they were designed for a certain purpose but there is no such design, and the purpose is a matter for synoptic judgment when considering populations of the organisms statistically.[21] If we take Darwin seriously, it is crucial that we do not ascribe intentionality or instrumental targeting to biological functions or we will mistakenly turn them into a system governed by the kind of instrumental rationality that we ascribe to our own intentional actions that are carried out "under description" (Anscombe 1963). This line of criticism is well-known from Fodor's criticism against teleosemantic theories of intentional content: "Darwin cares how many flies you eat, but not what description you eat them under" (Fodor 1990: 73).

The temptation to ascribe to biological functions a purpose is two-fold. First, running away from bears in the forest seems purposive because bears can kill us. So we can rightly say, the fear system looks like it has the purpose to avoid death by a bear, but it is only an 'as if' purpose. But this does not mean that we can qualify every purpose we wish with an 'as if', ascribe it to a biological function, and slip into re-introducing a new kind of blind design. The slippage begins when, as persons with concepts and instrumental rationality, we

[21] See Dennett's (1995: Ch. 21) understanding of evolutionary processes as mindless and mechanistic algorithms.

notice that bears are dangerous, and say that the fear system "as if" has the purpose to avoid dangers. This last claim is false.

The members of the category 'bear' all share a typical visual form (or any other sense modality) and typical motor programs that can plausibly be identified by a biological function as belonging to the same category that is in turn differentiated from other categories. Conversely, the members of the category 'danger' share very few attributes, normally describable by using equally abstract concepts. It is the kind of category that people with conceptual capacities and instrumental rationality have, not the kind that could plausibly be attributed to a biological function.

Arguably we could thin-down the concept 'danger' to a purely ecological category that would mean 'threat to life and limb'. But then the supposed tracking system would be very limited, excluding many if not most of the objects we fear in our everyday life, such as exams, public speaking, being late, being rejected in love or at work, etc. The concept 'danger' as we ordinarily apply it is normatively laden, and does not designate an ecological category.

A sub-personalist could perhaps forgo accounting for most of our fear reactions, limiting them to tracking "dangers" that are stripped of normativity. This is no small concession. But then another option suggests itself, that is, that our emotional system is *not* geared to track threats to life and limb but rather biologically basic *objects* and anything that is similar to or rather that imaginatively connects with them. I show elsewhere why the latter option is more explanatorily fruitful, but here I simply argue that the subset of scientific data invoked by sub-personalists significantly underdetermines their theory (Morag 2016).

In any case, the categories of the significant, extreme, clichéd, and biologically basic that in light of the experiments cited by sub-personalists may be said to include objects toward which we reliably emote in predictable and fitting ways *are not representative of most of our emotional life*. Life is not a soap opera or an advertisement. If it is like fiction, it is closer to the more risky, original and non-formulaic forms of story-telling (as in certain novels, films and TV Drama Series).

*Most* of our emotional lives take place in our *ordinary everyday circumstances*, at work, at home, at the café, in the supermarket, at dinner parties. Our emotions usually involve people and things we know—our friends, our colleagues, our roommates, our romantic partners, our family members, our neighbors, the barista, our pets, our stuff. Most of our emotional reactions do not involve the biologically basic (e.g. bears, being hit) or the clichéd (e.g. cute babies, football teams). Although there is a level of description in which romantic partners or family members are biologically basic objects, our quotidian emotional relationships with them involve much more than their sexual or care-giving or care-demanding functions. We react to them, rather, qua having shared experiences that are particular to our lives and qua having distinctive personalities. In fact, even the sexual aspect of our relationships is often idiosyncratic and goes far beyond any reproductive goal (see Freud 1905). The vast majority of our emotional life involves our *intimates*—people, animals and things, with whom we have ongoing and complex relationships, relationships that may go through occasional extreme upheaval or significant cross-roads, but that are emotionally maintained throughout ordinary and everyday situations. The experiments sub-personalists often cite thus do not provide good evidential support for the sensitivity or the accuracy of the supposed emotional tracking system, because they

only cover a relatively small portion of our emotional life. Indeed, it would be extremely challenging to collect enough data about people's everyday personal lives. Presumably, one on one psychotherapy over a few years is the only currently existing medium to study people's private lives, and even then the data is far from objective, as it is largely comprised of recounted memories, and the real-time emotional experiences are either addressed at people who are not present or at the therapist. The Strong and the Weak versions of the Tracking Dogma thus remain unsubstantiated by the philosophers of emotion that hold them.

## 3. The Empirical Non-Scientific Appeal to Ordinary Psychological Experience and Common Knowledge

So how can we know whether or not the emotions of the human adult by and large track core relational themes? How can we have epistemic access to most of our emotional reactions that take place not in a lab but in our personal everyday lives? Empirical knowledge that we have through experience need not be limited to the conclusions of scientific observation and experiments. Experience also includes *ordinary everyday experience*, the kind of experience that is commonly turned to in other areas of philosophy, even if without explicit discussion.

Consider that in moral philosophy we appeal to moral experience, in aesthetics we appeal to aesthetic experience, and in epistemology we appeal to perceptual experience. So, too, I want to claim in philosophy of psychology we can avail ourselves of *psychological experience*: that is, in addition to knowledge of our own minds,[22] our experience of the beliefs, desires, feelings and sensations of *other* people. In the philosophical literature this is often misleadingly called 'folk psychology' but understanding others cannot be assumed to take the form of a predictive scientific theory or a preliminary attempt at that. In many cases it is more a matter of trying to see things from another's point of view, to imaginatively stand in their shoes. But it also includes our capacity to "read" the meaning or significance of other people's actions, expressions, gestures, style, and so forth. And this, in turn, depends on what we have learnt about the human condition and its complicated modes of expression from how others interpret us and our social relations, from fairytales, art, novels, and also from old adages, proverbs, and aphorisms, which are paradigms of non-scientific modes of understanding. Psychologists attempt to provide a science of mind-reading, but our own ordinary mind-reading skills, assumed by our daily functioning and dependent upon other non-scientific practices, provide us with us with non-scientific empirical knowledge.

The idea that there is a category of non-scientific empirical knowledge is overlooked in contemporary philosophy given the wide popularity of scientific models of what there is and how we know it. We have forgotten that Hume, for example, relied on his own *subjective* experience of his own mental states, which cannot be the subject of scientific study since it does not meet the appropriate standards of objectivity such as impersonal and relatively definite standards of identifiability or verifiability. Such subjective experience can count as data for

[22] I note that Stanley Cavell (1979: 146) has complained that "the subject of self-knowledge […] as a source of philosophical knowledge has been blocked or denied in modern philosophy".

general statements about the psyche, inasmuch as one can appeal to one's own and others' subjective psychological experiences, based on the reasonable assumption that we all have the same basic psychological capacities. It is also worth recalling that ordinary language philosophers relied on their mastery of their own language, on their own experience of language, to make general claims qua one of many competent language speakers, and not qua linguist who studies language from a detached scientific perspective that links marks and noises (i.e. objects of scientific study) to certain behaviors (Cavell 1969: Ch. 1).

Similarly, each of us is in a good position to "read" and identify the emotional patterns of our intimates. We are in a position to know the emotional lives of the people closest to us, better than those who do not know them well or at all, and perhaps even better than themselves. Reading another's emotion requires more than identifying familiar facial expressions and other prototypical behaviors, which in any case people often successfully inhibit in the company of strangers. People express emotions in many idiosyncratic ways. When it comes to our intimates, we can tell how they feel by the way they say 'hello', by the way they place a glass on the dinner table, by many gestures that are typical to them in particular or by subtle departures from their normal personal style of behavior. Furthermore, although we know well only a limited number of people, we assume that our relationships are not so different to those of others, at least in the sense that we conduct our relationships under the pressure of shared social norms of language, culture and emotion fittingness. In other words, we have *common knowledge* about emotions, a familiarity with the emotional lives of ourselves and of others, on which I rely in the next section. It matters little that this knowledge is defeasible and fallible.

This appeal to common knowledge may go largely unrecognized in the contemporary literature, but it is not unknown in philosophy. Consider G.E. Moore's common sense claims such as that he knows that since his birth he has lived on or near the surface of the earth (Moore 1959: 33). Moore says this about himself, based on his own self-knowledge. He appeals to his readers to acknowledge that they know it too based on their own self-knowledge. This is a good example of an appeal to non-scientific empirical knowledge. Differently to Moore, I shall rely particularly on not my own experience but our common experience of *other* people, qua ordinary emoting subjects that are functioning members of a social niche guided by various familiar social norms, including norms of emotion-fittingness.

In fact, I implicitly relied on common knowledge in the previous section when endorsing the plausibility of the claim that we are sensitive to and emote fittingly in the face of the significant, the extreme, the clichéd, and the biologically basic. In what follows I mobilize common knowledge to list a number of common phenomena that do not sit well with the reliability claim. None of the phenomena I mention can refute it on its own. Indeed, clear empirical refutation or verification may be too demanding in the realm of emotion, where, to paraphrase what Aristotle says about ethics, we should not expect much precision (Aristotle 1991: 1094b 12-26). It is rather the cumulative weight of the phenomena listed in the next section that provides a reasonable doubt about the plausibility of sensitivity and accuracy of the supposed emotional tracking system. It is worth recalling that Peirce recommended this methodology for philosophy:

> Philosophy ought to imitate the successful sciences in its methods, so far as to proceed only from tangible premises which can be subjected to careful scrutiny, and to trust rather to the multitude and variety of its arguments than to the conclusiveness of any one. Its reasoning should not form a chain which is no stronger than its weakest link, but a cable whose fibres may be ever so slender, provided they are sufficiently numerous and intimately connected (Peirce 1868: 229).

## 4. Common Knowledge against Reliability of (Thematic) Tracking

### 4.1 Against Accuracy

1. *Ordinary Language Expressions*

A number of familiar expressions demonstrate that many of our emotional reactions do not serve to track anything that directly bears on well-being, but rather often seem counter-productive. That is, we think that people emote when they should not often enough to have expressions such as the advise to 'not take things personally', and gentle criticisms such as 'touchy!' or 'I guess I hit a nerve'. If such un-called for emotions tell us anything useful, it is about the emoters' "soft spots", as we say, about their issues and sensitivities. In fact, some of us accuse one another for taking advantage of those sensitivities and 'pushing our buttons'. And there is the familiar warning about a potential partner's 'emotional baggage'.

2. *Transference*

As Freud theorized, we often "transfer" and emote toward one person in a way that we would emote toward another person.[23] For example, a man may resent his boss for being domineering and hostile because she reminds him in some way of his mother, who he has been resenting for a long time for being domineering and hostile—whether or not his boss actually is domineering and hostile.[24]

One need not agree with every word Freud wrote or with various familiar psychoanalytic conceptions to see that transference is a commonly acknowledged and frequent enough phenomenon, as is attested by known expressions such as 'don't shoot the messenger', 'don't take it out on me', 'I am not your mother!', '*She's* having a good/bad day', and so on. In such cases we do not think people track anything with their emotion but rather expose to us their "emotional baggage" concerning other people and other (past and present) sets of circumstances. Whereas the Tracking Dogma identifies the here-and-now emotionally salient object of the emotion as its principal cause, experience often suggests that this here-and-now emotionally salient object is merely a causal trigger, and that there are other more significant causes and objects that are not even present, nor straightforwardly similar to the here-and-now circumstances. Whereas the Tracking Dogma assumes a purposive and instrumental relation between the emotion and the object it makes salient, the phenomenon of transference assumes an *imaginative* relation.

[23] Freud discovered the phenomenon of transference in the context of psychoanalytic therapy, when patients would emote toward the therapist in ways that are typical of their relationships with others in their lives. See for example Freud 1914.
[24] See the example of Jonah's resentment to his boss Esther in Rorty 1980.

3. *Projections*

Freud also spoke of "projections" whereby we ascribe to people (often falsely) qualities or emotions that we have ourselves. For example, people who entertain unfaithful thoughts often ascribe such thoughts or even actions to their romantic partners thereby suffering from what Freud called "projected jealousy" (Freud 1922: 224). In such cases, one's jealousy is not tracking any defection of affection. Such projected suspicions may give rise not just to jealousy, but also to anger and fear that do not track anything objective. The process of projection is familiar to many people who have never read Freud, and it has been made use of in novels before the term was coined. Just as an example, consider the old Hebrew saying that originates from the Talmud Bavli, written many centuries before Freud was born: "The fault one finds in another is one's own".[25] Here, once more, we see how imagination and emotion interconnect to make certain objects salient without it being instrumental or conducive to the subject's well-being. If transference-emotions or projective-emotions tell us anything at all—it is about the person's own subjective "soft spots."

4. *Practices of Emotion Inhibition*

Consider the prevalence of various strategies of self-management to dim down our emotions (but not through direct rational criticism of them). For example, we are told to count to ten before we express anger so that it will give us a moment to see if indeed the situation merits this anger.[26] Another familiar strategy to control unfitting emotions of all kinds consists in the recurrent advice or decision to 'just get over it' or 'stop thinking about it' and 'move on'. We learn, train ourselves, transmit and sustain social practices of controlling the expression of our emotions, at times by ignoring our emotion and focusing our attention on something else. Emotion inhibition is also exercised when emotions are fitting but otherwise socially unacceptable.[27] But at least some cases of emotion inhibition demonstrate our familiarity with the large scope of unfitting emotions.

5. *Psychotherapy*

The prevalence of unfitting emotions is explicitly acknowledged by the very existence of the practice of psychotherapy and the large number of people who seek psychoanalytic, psychological or psychiatric help to resolve emotional issues that appear to them to be out of kilter with reality. This provides *prima face* evidence of actual mismatch. In fact, psychoanalytic practice presupposes that emotions are not a rational phenomenon, and should not be judged as "fitting" or "unfitting". This is a claim about the practice rather than about specific theories one can find in the psychoanalytic literature—a claim that I defend elsewhere (Morag 2016: Ch. 6).

6. *Love*

'Love is in the eyes of the beholder'. 'Love is blind'. It is so well-known that love, especially romantic love but also friendship, cannot be judged as fitting or unfitting, that its object may often be not at all conducive to one's well-being,

[25] My translation of the known Hebrew saying: "Haposel Bemumo Posel".

[26] James: "Count ten before venting your anger, and its occasion seems ridiculous" (James 1983: 178).

[27] See Ekman and Friesen (1975: Ch. 2, 11). Display rules include not just rules of emotion inhibition but also of emotion exhibition.

and that it does not track anything that can be characterized in rough-and-ready terms, that most modern philosophers of emotion have simply excluded it from their list of emotions. Love may be painful, it may involve someone whose character is incompatible to ours, but such judgments are not expected to cause love to end. And yet, love is not merely a disposition to emote in various ways, as some philosophers claim (Roberts 1988). Love, in one important sense at least, is an emotion; it has an occurrent form with its own prototypical physiological and behavioral manifestations. We speak of having "butterflies" when we are in love. We can tell when someone is in love even when that person forcefully denies they are in love, for example, by the way they look at their love interest.[28]

7. *Moods*

Moods are by and large defined as affective states that do not make any particular object in the environment emotionally salient. The question of fittingness or of accuracy of tracking does not arise for them. Together with love, the overwhelming majority of philosophers of emotion have excluded moods from their accounts. And yet moods fit the minimalist definition of emotions presented in the beginning of this paper. They are characterized by longer episodes of the same types of other emotions: depression is like sadness, euphoria is like joy, irritability like anger, anxiety like fear. Moods may not make a specific object within the environment emotionally salient, but they do color the experience of one's environment as a whole with their affect. The world looks generally gloomy when we are depressed, or feels full of opportunities when we are optimistic. Saying that moods are not emotions since they do not make specific objects emotionally salient or because they do not have norms of fit and do not appear to track anything useful is an *ad hoc* claim, motivated by theoretical considerations. Why is it assumed that the emotionally salient object of a certain affective experience is its "intentional object" or "target"? Why could this object not simply be the causal trigger of the emotion or some other object in the nearby environment that the emoting subject is focusing on? Why must we assume that emotions are short-lived episodes? If we see moods as emotions, then their prevalence does not sit well with the idea that emotions reliably track purposive themes or that they track anything at all for that matter.

## 4.2 Against Sensitivity

If our emotional tracking system, said to be aimed at tracking dangers, wrongs, benefits, achievements, etc., were a sensitive system, then by and large most of the time we would respond in fitting ways to dangers, wrongs and other core relational themes. *And yet it is often the case that we can appreciate a situation as meriting an emotional response and fail to emote.* Let us call this the Emotionality Problem. Sometimes we are annoyed with drivers that cut in front of us on the road and sometimes not, even when our appreciation of their rudeness has not changed. Sometimes we may feel great sadness when we hear about the fighting in Syria and sometimes we hardly feel anything at all. Sometimes we jump in

[28] That love has an episodic nature with prototypical physiological "activation" is common knowledge but has also been recently proved scientifically (cf. Nummenmaa, Glerean, Hari & Hietanen 2014).

joy when a good old friend calls after a long period of absence and other times we remain indifferent.

Jenefer Robinson, after laying out her sub-personalist account, acknowledges the Emotionality Problem and the way it clashes with the Strong Tracking Dogma as follows: "Why am I emotional about something on some occasions and on other occasions not?" (Robinson 2005: 95). Robinson attempts to add other "variables" to the tracking system in order to account for this variability. As I show elsewhere (Morag 2016: Ch. 2, 3), Robinson's suggestions amount to relying on moods, whether these are caused by a certain physiological cause (e.g., hormones, drugs, fatigue, energy levels) or by an earlier event (e.g., confidence due to a promotion), or somehow otherwise caused.

Moods seem to indeed have an effect on our susceptibility to have shorter and more intense affective episodes that make specific people and things emotionally salient. Furthermore, often when people we know well emote in unfitting ways we ascribe to them such moods. We assume that the specific affect is already present, coloring the day's experiences accordingly. Some experiments show that manipulating aspects of physiology that are relevant to emotion, by forcing certain facial expressions (Ekman 1984: 324-28, cited in Robinson 2005: 36)[29] or levels of physical arousal through receiving drugs, can effect emotional experiences and the manner in which they are reported.[30]

But if moods were the only variable to explain the irregularity of our emotional reactions to otherwise fitting circumstances that bear on our relevant cares and concerns, then we would be obliged to attribute to ourselves moods every time we emote. This is *ad hoc*, and one would need a full account of moods to augment any account of emotions, to explain their emotionality. Whatever such an account may be, this would mean that emotions are constantly biased by our mood, thereby failing to track any purposive theme that does not fit an affective state that is not that mood. The idea that emotions track core relational themes would be effectively given up.

## 5. Alternatives to the Strong and Weak Versions of the Tracking Dogma

The above considerations cast serious doubt on the plausibility that the emotional system is sensitive to core relational themes or that whenever we do have an emotional reaction it accurately tracks dangers, wrongs, achievements, etc. Consequently I contend that the Strong and the Weak versions of the Tracking Dogma are implausible views. Two alternatives now suggest themselves.

The first is the Normative Tracking Dogma, a fallback position for those who still want to maintain that emotions aim to track purposive themes. Indeed, anyone who holds that emotions have intentional content is obliged to endorse this normative position, namely that emotions at least *aim* at fittingness, even if they often fail. Many philosophers who hold the Normative version of the Tracking Dogma are optimistic about our capacity to improve the accuracy our supposed tracking system, about our chances of becoming the Phronimos. It is

[29] See also Zajonc, Murphy & Inglehart 1989; as well as Strack, Martin & Stepper 1988 (cited in Prinz 2004: 35-36).

[30] Recall the famous Schachter and Singer (1962) experiment, discussed for example in Griffiths (1997: 25, 81), Prinz (2004: 71), Robinson (2005: 83-85).

what Karen Jones has named the contemporary "Pro-Emotion Consensus" according to which "emotions can, with experience and regulation, become reason-tracking mechanisms that enable an agent reliably to track the way her concerns are implicated in concrete choice situations" (Jones 2006: 4). But in light of the considerations presented in this paper, these philosophers should admit that the Aristotelian Phronimos who emotes fittingly is either extremely rare or non-existent. They can still claim that the Phronimos is a worthy ideal toward which we should all strive. To say it all too briefly, once we take seriously the variety and frequency of cases where emotions are not fitting to their circumstances, the Normative Tracking Dogma faces two main problems: 1) It assumes implausibly far-reaching irrationality in all adults, and 2) It lacks explanatory resources that would account for unfitting emotions.

The challenge for all views of emotions, a challenge that the Strong and Weak versions of the Tracking Dogma hardly admit let alone answer, is to explain how the relatively small pool of objects to which we reliably react in predictable and fitting ways in infancy develops into a much bigger pool of objects to which we *often* do not react in predictable or fitting ways. I conclude by briefly introducing the alternative vision and the specific account I favor and defend elsewhere as an account that can meet his challenge (Morag 2016).

Rejecting the Strong and the Weak versions of the Tracking Dogma already renders less appealing the idea that emotions *aim* at fittingness, that they are representational, that they have intentional content or some other form of information embedded in them. If this supposed representation is often mistaken, then perhaps the very idea of emotions as having content that either fits or does not fit the situation in which it arises is misguided. Emotional representationalism seems, in light of the considerations I have amassed here less compelling than the idea—defended most famously by Plato, Hume and William James in the philosophical tradition and taken for granted by the practice of psychoanalysis—that emotions are not representational and are not rationally assessable in terms of "fit".[31]

According to my version of this non-representationalist vision, emotions should not be seen as either succeeding or failing to be revelatory about the world, as the Tracking Dogma takes them to be. Rather, emotions are revelatory of the inner life of the mind that embeds one's past experiences. To summarize all too briefly an account I present elsewhere,[32] I propose what we may call an imagistic seeing-as account, whereby I see here-and-now people and things in terms of other people and things that were emotionally salient in the past, and not in terms of concepts such as danger or wrong. That is, when I imaginatively connect a here-and-now "object" (through similarity, inversion or part-whole relations) to past "objects" of past or remembered emotional experiences, the here-and-now "object" becomes emotionally salient. According to this alternative then, all emotions are transference or projection emotions, and their fitting-

[31] Hume writes: "A passion is an original existence, or, if you will, modification of existence, and contains not any representative quality, which renders it a copy of any other existence or modification" (Hume 1738: Book II, Part 3, Section 3). I defend the (controversial) claim that psychoanalytic practice assumes a non-representational view in Morag 2016: Ch. 6.

[32] See Morag 2016: Part 2 for my proposed positive account for the formation and subsidence of emotional reactions.

ness to the here-and-now is a contingent matter of an after-the-fact normative judgment. Emotions, in other words, do not detect how the world objectively relates to us and our well-being, but rather express our subjective and personal way of seeing the world through the imaginative lens of our experiences and memories.[33]

## References

Anscombe, E. 1963, *Intention,* Oxford: Basil Blackwell, 1985.

Aristotle 1991, *Nicomachean Ethics*, in *The Complete Works of Aristotle*, Vol. 2, Barnes, J. (ed.), Princeton: Princeton University Press.

Baier, A. 2004, "Feelings that Matter", in Solomon, R. (ed.), *Thinking about Feeling*, New York: Oxford University Press, 200-13.

Brady, M. 2007, "Recalcitrant Emotions and Visual Illusions", *American Philosophical Quarterly*, 44, 3, 273-84.

Cavell, S. 1969, *Must We Mean What We Say?*, Cambridge: Cambridge University Press.

Cavell, S. 1979*, The Claim of Reason*, Oxford: Oxford University Press.

Charland, L. 1995, "Feeling and Representing: Computational Theory and the Modularity of Affect, *Synthese*, 105, 3, 273-301.

D'Arms, J. 2000, "Empathy and Evaluative Inquiry," *Chicago-Kent Law Review*: *Symposium on Law, Psychology and the Emotions*, 74:4, 1467-1500.

D'Arms, J. and Jacobson, D. 2003, "The Significance of Recalcitrant Emotion (or, Anti-Quasijudgmentalism)", in Hatzimoysis, A. (ed.), *Philosophy and the Emotions,* Cambridge: Cambridge University Press, 127-46.

De Sousa, R. 1987, *The Rationality of Emotion*, Cambridge (MA): MIT Press, 1990.

Debes, R. 2009, "Neither Here nor There: The Cognitive Nature of Emotion", *Philosophical Studies*, 146, 1, 1-27.

Dennett, D. 1995, *Darwin's Dangerous Idea: Evolution and the Meaning of Life,* New York: Simon and Schuster.

Deonna, J.A. 2006, "Emotion, Perception and Perspective," *Dialectica*, 60, 1, 29-46.

Döring, S. 2003, "Explaining Action by Emotion," *The Philosophical Quarterly,* 53, 211, 214-30.

Döring, S. 2010, "Why Be Emotional?", in Goldie, P. (ed.), *The Oxford Handbook of Philosophy of Emotion*, Oxford: Oxford University Press, 283-302.

Ekman, P. 1984, "Expression and the Nature of Emotion", in Scherer, K.R. and Ekman, P. (eds.), *Approaches to Emotion*, Hillsdale (NJ): Lawrence Erlbaum Associates, 319-44.

Ekman, P. and Friesen, W. 1975, *Unmasking the Face: A Guide to Recognizing Emotions from Facial Expressions*, Cambridge (MA): Maylor Books, 2003.

Field, T. *et al.* 1982, "Discrimination and Imitation of Facial Expressions by Neonates", *Science*, 218, 179-81.

Fodor, J. 1990, *A Theory of Content and Other Essays*, Cambridge (MA): MIT Press.

[33] I am very grateful to the blind referees and especially to David Macarthur for the comments on earlier versions of this paper.

Freud, S. 1905, *Three Essays on the Theory of Sexuality*, S.E., Vol. 7, 123-243.

Freud, S. 1914, "Remembering, Repeating, and Working-Through", S.E., Vol. 12, 147-56.

Freud, S. 1922, "Some Neurotic Mechanisms in Jealousy, Paranoia and Homosexuality", S.E., Vol. 18, 223-32.

Friesen, W.V. 1972, *Cultural Differences in Facial Expressions in a Social Situation: An Experimental Test of the Concept of Display Rules*, Doctoral dissertation, San Francisco: University of California.

Frijda, N. 1986, *The Emotions,* Cambridge: Cambridge University Press.

Greenspan, P. 1988, *Emotions and Reasons,* London: Routledge.

Griffiths, P. 1997, *What Emotions Really Are: The Problem of Psychological Categories,* Chicago: Chicago University Press, 1998.

Griffiths, P. 2004a, "Is Emotion a Natural Kind?", in Solomon, R. (ed.), *Thinking about Feeling: Contemporary Philosophers on Emotion*, Oxford-New York: Oxford University Press, 233-49.

Griffiths, P. 2004b, "Towards a Machiavellian theory of Emotional Appraisal", in Evans, D. and Cruse, P. (eds.), *Emotion, Evolution and Rationality*, Oxford: Oxford University Press, 89-105.

Hume, D. 1738, *A Treatise of Human Nature*, Selby-Bigge, L.A. (ed.), Oxford: Clarendon Press, 1978.

Izard, C.E. 1978, "On the Ontogenesis of Emotions and Emotion-Cognition Relationship in Infancy", in Lewis, M. and Rosenblum, L. (eds.), *The Development of Affect*, New York: Plenum Press, 389-413.

James, W. 1983, "What is an Emotion?", in Burkhardt, F. *et al*. (eds.), *The Works of William James*, Vol. 13, *Essays in Psychology*, Cambridge (MA): Harvard University Press, 1983, 168-87.

Jones, K. 2006, "Quick and Smart? Modularity and the Pro-Emotion Consensus," *Canadian Journal of Philosophy*, 36, Suppl. Vol. 32, 3-27.

Lazarus, R.S. 1991, *Emotion and Adaptation*, New York: Oxford University Press.

Lazarus, R.S. and Alfert, E. 1964, "Short-Circuiting of Threat by Experimentally Altering Cognitive Appraisal", *Journal of Abnormal and Social Psychology*, 69, 195-205.

Logue, A.W., Ophir, I. and Strauss, K.E. 1986, "Acquisition of Taste Aversions in Humans", *Behavioural Research and Therapy*, 19, 319-33.

Lyons, W. 1980, *Emotion*, Cambridge: Cambridge University Press.

Maranon, G. 1924, "Contribution a l'etude de l'action emotive de l'adrenaline", in *Revue Francaise d'Endocrinologie*, 2, 301-325.

Meltzoff, A.N. and Moore, M.K. 1977, "Imitation of Facial and Manual Gestures by Human Neonates", *Science*, 198, 75-78.

Meltzoff, A.N., and Moore, M.K. 1989, "Imitation in Newborn Infants: Exploring the Range of Gestures Imitated and the Underlying Mechanisms", *Developmental Psychology*, 25, 954-62.

Meltzoff, A.N. and Moore, M.K. 1995, "Infants' Understanding of People and Things: From Body Imitation to Folk Psychology", in Bermudez, J., Marcel, A. and Eilan, N. (eds.), *The Body and the Self*, Cambridge (MA): MIT Press.

Moore, G.E. 1959, "A Defence of Common Sense", in *Philosophical Papers*, London: Allen & Unwin, 32-45.

Morag, T. 2016, *Emotion, Imagination, and the Limits of Reason,* London: Routledge.

Nummenmaa, L., Glerean, E., Hari, R. and Hietanen, J.K. 2014, "Bodily Maps of Emotions", *Proceedings of the National Academy of Sciences of the United States of America*, 111, 646-51.

Ohman, A., Fredrikson, M. and Hugdahl, K. 1976, "Premiss of Equipotentiality in Human Classical Conditioning", *Journal of Experimental Psychology*, 105, 313-37.

Peirce, C.S. 1868, "Some Consequences of Four Incapacities", in Buchler, J. (ed.), *Philosophical Writings of Peirce*, New York: Dover, 1955, 228-50.

Prinz, J. 2004, *Gut Reactions: A Perceptual Theory of Emotion,* New York: Oxford University Press.

Roberts, R. 1988, "What an Emotion Is: A Sketch", *The Philosophical Review*, 97, 2, 183-209.

Robinson, J. 2005, *Deeper than Reason*: *Emotion and its Role in Literature, Music and Art*, Oxford: Clarendon Press.

Rorty, A. 1980, "Explaining Emotions", in Rorty, A. (ed.), *Explaining Emotions*, Berkeley: University of California Press, 103-26.

Salmela, M. 2011, "Can Emotion be Modeled on Perception?", *Dialectica*, 65, 1, 1-29.

Schachter, S. and Singer, C. 1962, "Cognitive, Social, and Physiological Determinants of Emotional State," *Psychological Review*, 69, 379-99.

Shin, L.M., Dougherty, D.D., Orr, S.P., Pitman, R.K., Lasko, M., Macklin, M.L., Alpert, N.M., Fischman, A.J. and Rauch, S.L. 2000, "Activation of Anterior Paralimbic Structures During Guilt-Related Script-Driven Imagery", *Biological Psychiatry*, 48, 43-50.

Sroufe, A.L. 1984, "The Organization of Emotional Development", in Scherer, K.R. and Ekman, P. (eds.), *Approaches to Emotion*, Hillsdale (NJ): Lawrence Erlbaum Associates, 109-28.

Strack, F., Martin, L.L. and Stepper, S. 1988, "Inhibiting and Facilitating Conditions of Facial Expressions: A Nonobstrusive Test of the Facial Feedback Hypothesis", *Journal of Personality and Social Psychology,* 54, 768-77.

Tappolet, C. 2012, "Emotions, Perceptions, and Emotional Illusions", in Calabi, C. (ed.), *Perceptual Illusions: Philosophical and Psychological Essays,* London: Palgrave Macmillan, 207-24.

Trevarthen, C. 1984, "Emotions in Infancy: Regulators of Contact and Relationship with Persons", in Scherer, K.R. and Ekman, P. (eds.), *Approaches to Emotion*, Hillsdale (NJ): Lawrence Erlbaum Associates, 129-57.

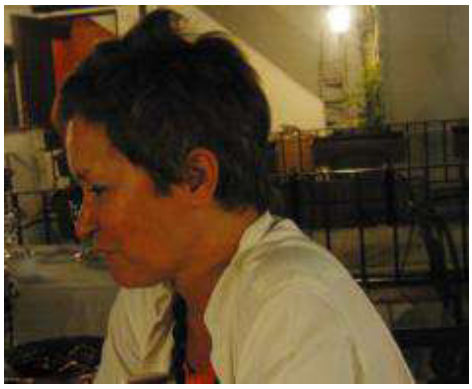
Watson, J.B. 1924, *Psychology from the Standpoint of a Behaviorist*, 2nd edition, Philadelphia: Lippincott.

Zajonc, R.B., Murphy, S.R. and Inglehart, M. 1989, "Feeling and Facial Efference: Implications of the Vascular Theory of Emotion, *Psychological Review*, 96, 395-416.

# The Style of Philosophy: An Obituary for Eva Picardi

*Paolo Leonardi*

*University of Bologna*



On Sunday afternoon, April 23, 2017, Eva Picardi died after a long illness. Eva had been Professor of Philosophy of Language at the University of Bologna for forty years. She belonged to a small group of Italians of her generation who did not just study and discuss contemporary analytic philosophy but was herself an active member of that larger philosophical community.

Eva had style—philosophical and personal. She mastered her field and had knowledge beyond it. Eva was no sceptic, and had firm philosophical certainties—she was a Fregean and worked on anything that is problematic in Frege's philosophy of language and its aftermath. In discussion, she was precise and insightful. At the same time, often she did not argue the last steps: references and quotations insinuated a different ground and the unfinished argument left open the conclusion. It was lightness and respect, and more. She was as convinced that matters can be seen more than one way, which is what rewards us in a vast knowledge of the literature. A perspicuous picture, which is what we correctly aim at, is one that looks at its object from each of the surrounding points for an indefinite span of time, i.e. an impossible picture. That is no cause for regret—the world and life are richer than any pictures of them. We rather make maps, which takes notice of the asperities on the grounds, of the traffic there, of where we get supplies, with occasional glosses of phantasies we can make in traveling there. The maps help us explore spaces and habits that change through time more or less dramatically, requiring us to update them continuously.

*Argumenta* 2, 2 (2017): 365-369
ISSN 2465-2334


DOI 10.23811/411.arg2017.leo

A great teacher, Eva motivated and directed her students, a professional undertaking to which she dedicated time and energy, organizing seminars and students' presentations lasting a few whole days, during which she publicly discussed arguments almost line by line. At least ten of her students have become researchers and professors in Italy (Bologna, Pavia, two in Milan), other European countries (Birmingham, Helsinki, Lisbon), New Zealand (Auckland), US (Irvine, Northwestern University). As a member of several editorial boards, she acted as a referee, also for *Argumenta*, till the last months of her life, as she did working with Carlo Penco at a new translation into Italian of Frege's works.

Besides Frege—whose views of assertion were the topic of her doctoral work at Oxford, with Michael Dummett as supervisor—Wittgenstein was her author, whom too she studied and examined in depth. But Eva considered ideas of many other contemporary analytic philosophers: early classics, especially Russell and Ramsey, then Quine, Davidson, and more occasionally Brandom, and many of her generation or younger like Putnam, Travis, Soames, Lepore. As well she studied nineteenth-century logicians, and examined differently oriented philosophers like Peirce, Husserl, Gadamer, etc. I have met three philosophers with a full knowledge of their field, its whole literature and with a precise memory of their readings, which they could quote by heart—Eva was one of the three. Like the other two, she knew much more than her philosophical province and was curious about science, literature, music too.

Just as I cannot list the authors Eva studied in the course of ninety papers and four books, I cannot enumerate the topics she took up. Together with Annalisa Coliva, she edited a book of essays on Wittgenstein, *Wittgenstein Today* (2004). Her papers are about topics such as vacuous names, radical interpretation, truth theories, compositionality, meaning and rules, belief and rationality, first-person authority, naturalism, sensory evidence, reference, normativity and meaning, grasping thoughts, concept and inference, literal meaning, multiple propositions, and identity. Eva's main contributions were the books *Assertibility and Truth. A Study of Fregean Themes* (1981), on the same topic as her Oxford D.Phil., and *La chimica dei concetti* (1994), which collected, in revised form, some of her essays on logic and psychologism, Frege and Kerry, Frege, Russell, Wittgenstein and Ramsey. She also published two introductions to the philosophy of language, *Linguaggio e analisi filosofica* (1992), *Le teorie del significato* (1999).

Among the many pieces, some are very brilliant, such as the "First Person Authority and Radical Interpretation" (1993), which criticizes Davidson's understanding of belief. With style, as if were an euphemism, and referring back to Frege's semantics, she entitled her essay on expressive content *Colouring*. All the papers she collected in *La chimica dei concetti* (*The Chemistry of Concepts*) are excellent. The title essay—whose ancestor, like those of other three papers there, was originally published in English with the same title—is about Frege's use of the metaphor of unsaturatedness, in close comparison with Peirce's use of it. The first essay in the collection, which too had an English ancestor, "The Logic of Frege's Contemporaries" (1987), maps late nineteenth-century German logicians, philosophers and psychologists, well- and less-well-known, whom Frege quoted or who quoted him, who in the then-dominant psychologistic approach to logic had ideas similar to some of Frege's. Sometimes the hints are captivating, as when Eva writes that to grasp what Lotze or Helmholtz meant respectively by *Lokalzeichen* (local sign) and by *Zeichen* (sign) "we have to look at their discussions of visual and spatial perception" (1987: 176). If Frege viewed lan-

guage as capable of overcoming the psychological conditioning of linguistic symbols making conceptual thinking possible (see 1987: 184), he, comments Eva, could not dismiss the fact that grasping thoughts, judging and presenting them as holding good "are mental acts occurring in time and performed by actual people" (1987: 186). That is, he could not dismiss understanding. (One of many her critical remarks of Frege's work.) The paper shows the vast knowledge Eva had of the history of logic in the years that Frege was at work. The essay discusses Frege and Husserl, Frege and Erdmann, rebuts the idea that Frege was a Kantian, analyses the convergences between Sigwart and Frege and of those between Wundt and Frege. For instance, she relates Wundt's view of judgment as a whole from which concepts are extracted to Frege's idea that concepts occur instantiated and "concept words are extracted from the complete sentences in which they occur" (1987: 193).

The paper on "Sense and Meaning" (which is a more faithful translation of the German *Sinn und Bedeutung*) is a reconstruction of the two notions of Frege's philosophy of language from 1879 until his last writings. Eva lists some problems of the distinction which she nevertheless endorses: (a) it is difficult to keep sense apart from meaning in the case of many expressions, such as connectives, prepositions, features, and numerals. (b) Linguistic signs are not signs of something, but *conventional* signs. Just how linguistic conventions work is, she stresses, a source of many misconceptions. (c) The context principle and the compositionality principle act differently for senses and for meanings: the sense of a complex expression composes the senses of the component expressions, the meaning of a complex expression, on the other hand, is a function of the meanings of simpler ones. (d) The assimilation of sentences to proper names—proper names of the True or of the False—deprives sentences of their central role in the context principle. (e) What do we learn when we come to know that an identity statement of the form $a=b$, which is not a priori, is true? Do we learn something about the signs, their sense or the object they mean? Eva answers the question by observing that the different senses of the names and the descriptions occurring in the identity judgment account for its informative value and are not what the identity judgment is about.

"Wittgenstein and Frege on Proper Names and the Context Principle" (2010), which had a German ancestor that Eva published in 2009 in *Deutsche Zeitschrift für Philosophie*, is first a comparison between Frege's *Grundlagen* on the context principle and Wittgenstein's *Tractatus* views on sense and meaning, then a comparison between Frege's *Grundlagen* on the context principle and Wittgenstein's *Investigations* language-game. The last section of the essay defends a Fregean Wittgenstein on proper names against Kripke's criticism. The *Investigation*'s view of proper names is not antagonistic to Frege's. Vacuous names, such as *Nothung*, proper names of ordinary people, like *N.N.*, proper names of famous persons like *Moses* are discussed in the book, and the focus is how we understand these proper names, a very Fregean theme. The concept of a proper name is a family resemblance concept, writes Eva, an idea that finds support in Peter Geach, who in Geach 1969 argues that a proper name cannot be tied to just one definite description (cf. Picardi 2010: 181-82) In Frege's footnote on the name *Aristotle* (in "On Sense and Meaning") or in his remarks on the name *Gustav Lauben* (in "The Thought") "what matters are the (fluctuating) pieces of information that can help speaker and hearer to find out whom they are talking about […] by using the same proper name" (Picardi 2010: 182). As in the case of

*Moses*, which is a name one uses without defining it by a fixed definite description. A case which is very similar to that of *Gustav Lauben* in Frege's "The Thought". The paper, as is typical of Eva, offers a classic point of view for contemporary disputes, here those of Charles Travis, François Recanati, Ernie Lepore and Herman Cappelen.

The precision and the care in reconstructing Frege and updating his philosophy of language came out, in 2002, when Saul Kripke spent just over a month lecturing in Bologna on Frege on sense and reference. Traces of their discussion, which went on even after class, can be gleaned from Saul's acknowledgments in the footnotes of Kripke 2008.

Eva's impact on the Italian philosophical scene was also due to her work as a translator and editor, and as President of the Italian Society for Analytic Philosophy 2000-2002. She translated into Italian and edited *The Posthumous Works* by Gottlob Frege; *Inquiry into Meaning and Truth* and *Essays on Actions and Events* by Donald Davidson; *Origins of Analytic Philosophy*, *The Logical Basis of Metaphysics*, *Thought and Reality* and *The Nature and Future of Philosophy* by Michael Dummett. Besides, she edited the Italian editions of *Thought and Reality* by Michael Dummett and of *The Threefold Cord. Mind, Body, and World* by Hilary Putnam. She translated into Italian *Wittgenstein's Lectures on the Foundations of Mathematics, Cambridge 1939*. With Carlo Penco she edited a collection of philosophical essays written by Frege between 1891 and 1897, including the three classic 1891-1892 papers. Besides, together with Joachim Schulte, she edited the German edition of *Inquiry into Meaning and Truth* by Donald Davidson.

During her long illness, she neither hid her condition nor turned it into a problem to share, going on as if there were no deadlines, undertaking various new projects. Just a few weeks ago, she was still making plans for Summer 2017. Elegant and beautiful, intelligent and cultivated, fearless.

If you want to see Eva's serious irony, have a look at this video: https://vimeo.com/108575306 (at 44 min, 11 sec; and 1 h, 19 min, 37 sec).

If you want to listen to a lecture by Eva, look this other video: http://www.cattedrarosmini.org/site/view/view.php?cmd=view&id=213&menu1=m2&menu2=m37&menu3=m410&videoid=935

## References

Geach, P.T. 1969, "The Perils of Pauline", *Review of Metaphysics*, 23, 287-300.

Kripke, S. 2008, "Frege's Theory of Sense and Reference: Some Exegetical Notes", *Theoria*, 74, 181-218.

Picardi, E. 1981, *Assertibility and Truth. A Study of Fregean Themes*, Bologna: CLUEB.

Picardi, E. 1987, "The Logics of Frege's Contemporaries or 'der verderbliche Einbruch der Psychologie in die Logik'", in Buzzetti, D. and Ferriani, M. (eds.), *Speculative Grammar, Universal Grammar and Philosophical Analysis of Language*, Amsterdam: Benjamins, 173-204.

Picardi, E. 1992, *Linguaggio e analisi filosofica. Elementi di filosofia del linguaggio*, Bologna: Pàtron.

Picardi, E. 1993, "First-Person Authority and Radical Interpretation", in Stoecker, R. (ed.), *Reflecting Davidson. Donald Davidson Responding to an International Forum of Philosophers*, Hawthorne: De Gruyter, 197-209.

Picardi, E. 1994, "Senso e significato", in *La chimica dei concetti*, Bologna: il Mulino, 109-80.

Picardi, E. 1999, *Le teorie del significato*, Roma-Bari: Laterza; Spanish Transl. Madrid: Alianza Editorial, 2001.

Picardi, E. 2006, "Colouring, Multiple Propositions, and Assertoric Content", *Grazer Philosophische Studien*, 72, 49-71.

Picardi, E. 2010, "Wittgenstein and Frege on Proper Names and the Context Principle", in Marconi, D., Frascolla, P. and Voltolini, A. (eds.), *Wittgenstein: Mind, Meaning and Metaphilosophy*, London: Palgrave, 166-87.

Balint, Peter, *Respecting Toleration. Traditional Liberalism and Contemporary Diversity*.
Oxford: Oxford University Press, 2017, pp. viii + 167.

After decades of neglect, the concept of toleration has become a central concern for moral and political theory since the late eighties when the issues arising from contemporary pluralism and cultural differences became paramount. The renewed interest for toleration, for exploring the possibility of such a concept as a tool for civil coexistence of differences in liberal democracy, has been however surrounded by doubts about its limits. From a political point of view, toleration has long been denounced as the counterfeit of despotism,[1] or as a disguised form of state repression.[2] From a social point of view, already Goethe remarked that no one likes to be tolerated, for what we want is recognition.[3]

These issues give rise to the three challenges to toleration from which Peter Balint's argument moves: the multicultural, the despotism and the neutrality challenge. The *multicultural challenge* in a way rephrases Goethe's objection to toleration, namely that it is only a second best, and does not grant inclusion to members of minority groups on an equal basis as members of majority. The *despotism challenge* instead rephrases Paine's concerns with toleration as a discretionary practice in the hands of political authority. Lastly, the *neutrality challenge* can be seen as a mixture of the criticisms both by Paine and Marcuse, for, if the state is neutral as to religion and ways of life, then toleration is redundant, and works as a repressive form of homogenization of differences. Taken together, these challenges lead to two opposite claims in the current discussion on toleration: the first according to which toleration is redundant, given liberal neutrality, the second according to which, instead, it is insufficient for minority differences to be equally included.

Balint's work takes issue with these two claims and presents a thorough defense of liberal toleration, considered as a freedom maximizing practice in the context of contemporary diversity. Balint offers a revised conception of toleration focused on the outcome of enhancing individual liberty in the face of diversity and of promoting a society where people having different and often incompatible ways of life are free to live as they see fit with minimal social and political restrictions. The emphasis on outcomes rather than attitudes or reasons is the first point of departure of Balint's view from the current discussion; this leads him to understand toleration as comprising two complementary views, a general permissive view of toleration and a more traditional conception of forbearance tolerance. Taken together, and applied at the state as well as at the citizens' level, these two understandings show that traditional liberalism, focused on negative liberty, has resources both to critique existing institutions for failing to properly apply neutrality and to accommodate the widest range of diversity in the same society. This is the core of Balint's original argument on toleration which articulates in three different steps: first, a critical stand against the mainstream view of toleration; second, a purely descriptive understanding of the con-

[1] Cf. Paine, T. 1989 [1791], "The Rights of Man", in Kuklick, B. (ed.), *Political Writings*, Cambridge: Cambridge University Press, 83-332 (94).
[2] Cf. Marcuse, H. 1969, "Repressive Tolerance", in Wolff, R.P., Moore, B. and Marcuse, H. (eds.), *A Critique of Pure Tolerance*, Boston: Beacon Press.
[3] Goethe, J.W. 1998, *Maxims and Reflections*, Hutchison, P. (ed.), London: Penguin.

*Argumenta* 2, 2 (2017): 371-383
ISSN 2465-2334

DOI 10.23811/412.arg2017.rev

cept of toleration against the prevalent moralized conception; third, a rebuttal of the expansion of toleration towards recognition and respect for differences. All three steps are meant to enhance the practice of toleration where what counts is behavior and outcomes instead of the right attitudes and reasons, in a context of diversity understood as different preferences rather than different cultures, religions, views.

The first step proposes a view of tolerance in line with the commonsense understanding of the tolerant society as one where people can live freely despite their different views and preferences. Contrary to the standard concept of toleration, implying the three components of objection, power to interfere, and withholding of the interference in favor of toleration,[4] a society is defined tolerant as far as it is permissive, without requiring a corresponding amount of objection and disapproval. In other words, Balint affirms that "being tolerant" refers to the wide range of activities and differences permitted in that society and not to the level of forbearance, for usually in a tolerant society there are fewer reasons to forebear, given that there are less reasons to object in the first place. In this general sense, indifference rather than objection is a relevant condition of permissive tolerance. For toleration as a general practice to be the case, only the power condition is relevant, meaning that the tolerator must have the effective power to hinder the different behavior or practice under consideration and nevertheless refrain to use the power of negative interference. The general practice of toleration does not rule out toleration as forbearance which is still needed in many cases. In case of forbearance tolerance, the three conditions of the standards model apply: 1. objection to a certain behavior, 2. power to negatively interfering with that behavior, and 3. withholding with that power. This general model is interpreted by Balint as purely descriptive in contrast with the prevalent tendency of philosophers to moralize toleration, either specifying that the objection must be of moral character or sustained by moral reasons or that the reasons to withhold the power of negative interference be of moral nature.[5] Balint is right in judging the moralized conception of toleration as unduly narrow, and mostly inapplicable in all interesting cases of contemporary toleration. The moralized conception is favored for a rigorous definition of toleration as a virtue, and also because it poses interesting philosophical puzzles as the case of the tolerant racist. If the good of toleration consists in sacrificing one's moral convictions for the sake of higher moral principles such as individual autonomy, authenticity and respect, then the stronger the objection, the more valuable toleration is, with the paradoxical result that a racist not acting on his racist conviction turns out more virtuous of a non-racist who does not have to overcome any objection to racial coexistence. Instead of devising ways out of the paradox, Balint gives up the moralized conception, and presents his descriptive forbearance tolerance without restrictions on the nature of the objections or on the reasons for non-hindrance. The combined result of the two steps considered so far is an enlarged view of toleration, where any act or omission by agents endowed with the pow-

[4] Cf. King, P. 1976, *Toleration*, New York: St. Martin's Press; Forst, R. 2003, "Toleration, Justice and Reason", in McKinnon, C. and Castiglioni, D. (eds.), *The Culture of Toleration in Diverse Society. Reasonable Tolerance*, Manchester: Manchester University Press, 71-85; Galeotti, A.E. 2015, "The Range of Toleration: From Toleration as Recognition back to Disrespectful Tolerance", *Philosophy and Social Criticism*, 41, 2, 93-110.

[5] Cf. Forst, R. 2013, *Toleration in Conflict*, Cambridge: Cambridge University Press.

er of negative interference, but refraining to use it, is comprised within the range of toleration, whether caused by forbearance, indifference, respect for difference. The final step of Balint's argument is to show that toleration in his liberal understanding is the best way to accommodate differences, without being too demanding on citizens as it were the case with the claim that differences ought to be respected, and producing better results, given that, in his view, respect for differences and difference accommodation are two distinct things not necessarily reconciled with each other.

Balint's defense of toleration is refreshing after so many attacks on either the insufficiency or the excess of toleration. I think he is right in considering toleration a crucial tool for dealing with contemporary diversity. I also agree on his view that neutrality is not alternative to toleration but rather is its embodiment within liberal state, though his argument is different from mine. On the whole, however, his work presents significant problematic aspects. A crucial general problem lies in Balint's outcome-oriented approach to toleration which, on the one hand, enlarges the scope of toleration but, on the other, flattens toleration on liberty, and actually makes it redundant. Toleration in fact *consists* in letting people free to pursue whatever ideals or way of life they prefer. Thus any toleration act is an omission of negative interference, and non-interference is precisely what negative liberty consists in. So what toleration adds to negative liberty and to the political and moral duty to respect others' liberty? The specific role of toleration lies precisely in both *accounting* and *providing reasons* for the acts of agents who are confronted with behavior or practice they dislike and have the power to hinder, but who decide not to use their power. Toleration, on the one hand, *explains why* agents refrain to use the power at their disposal to hinder behavior they dislike, and, at the same time, *provide reasons* for agent's self-restraint. The motivational component is in this sense crucial. From a purely behavioral point of view, the tolerant agent simply respects others' liberty. But the specificity of toleration lies in that the tolerant agent is confronting something that she does not like and that she has power to interfere negatively, but *choose* not to. So the point of toleration is precisely to understand how that is possible, and consequently to analyze the *attitudes and the reasons for self-restraint.* If there is *no need* to self-restraint, then toleration is utterly superfluous, and a purely behavioral account simply misses the original problem for which toleration has represented a solution, for the alternative of toleration is conflict. In this context, I wonder why if outcome and behavior are all that matters for Balint's toleration, he is insisting on the distinction between tolerance and endurance (116).

From this remark, it also follows that the purely descriptive view by Balint is unconvincing. Though I share his criticism towards moralized views of toleration, it does not follow that a non-moralized view of toleration should be purely descriptive. After all, toleration is a political ideal, is something that we consider on the whole good and definitely preferable to intolerance, as Balint concedes. A conception of toleration should obviously provide the conditions for toleration to be the case and these conditions are also descriptive. But the framework within which such conditions are placed is normative, even if leaving open the reasons why toleration is a good thing, be it of pragmatic, strategic or moral nature. The problem with a purely descriptive and behavioral conception of tolerance is that no boundary to toleration is considered central to the concept; boundaries in fact make sense only to circumscribe the area within which toleration is a good thing from the area where it turns into culpable indulgence. Yet

boundaries are crucial for toleration and for contemporary disputes around toleration. Even if we understand the expression "zero-tolerance of crime", it is deeply misleading to consider toleration of headscarf and tolerance of rape on a par. That toleration as a value has boundary is not a secondary feature of a moralized view, but inscribed in the traditional doctrine since Locke, and it comes down to the two principles of self-defense of the political order and harm to third party. Though their interpretation is controversial, nevertheless the two principles are pretty straightforward and generally accepted for setting apart a tolerant instance from complicity in crime. Thus Balint's example of the neighbor who becomes aware of domestic violence and does nothing, for whatever reasons, is not an example of tolerance, but of culpable indulgence. In order to set apart such culpable indulgence from tolerance of headscarf descriptively, we need to mark the scope of toleration normatively with the two principles setting limits to toleration as a value. Such limits correspond to the practice and the common usages of toleration, and divide what is tolerable from what is intolerable and must be prosecuted, such as rape, homicide, assault. If a too moralized conception of toleration unduly restricts its scope, a purely descriptive concept is too unrestricted and loses its specificity. Thus we do not need either for making sense of contemporary issues of toleration. Despite his intention, Balint himself cannot keep his presentation completely descriptive, for here and there he makes implicit reference to the harm principle, as when he says that toleration is simply "allowing people to do the *non-harming* thing they want to do/be" (88) and then he speaks of "acts of *unjustified* intolerance" (88), or "what matters is that if an individual finds himself in a situation where they could *unjustifiably* hinder another, they do not do so" (97).

A third point of Balint's argument which I find unclear is his reference to "political toleration", by which he does not mean only toleration by the state or public authorities or officials, but also toleration among citizens, yet considered not in their political dimension but as social agents. Basically he excludes from political toleration only "interpersonal toleration". In his view political does not coincides with vertical, while horizontal is not necessarily social. But in which sense social issues of toleration between groups are political is not clear to me, for he rules out that they are political because the political authority is ultimately the arbiter in cases of toleration conflicts which do not find spontaneous settlement. This I take is the typical circumstance of toleration issues in contemporary society where a horizontal dislike causes a conflict that requires a vertical decision. In sum, I think that by "political toleration" he means relevant cases of toleration which reach national media and public forum.

A final remark on the problem of respect which Balint takes up in Ch. 5 criticizing the thesis according to which respect for differences should replace toleration as the proper way of dealing with contemporary diversity. His argument is that respect for differences is too demanding on citizens while it is not necessarily conducive to difference accommodation. My criticism does not concern the thesis but the argument, which relies on a questionable interpretation of Darwall's recognition-respect.[6] In line with his general behavioral and anti-attitudinal approach, recognition-respect is seen as the outward behavior ac-

[6] Darwall, S. 1977, "Two Kinds of Respect", *Ethics*, 88, 39-49; 2006, *Respect and the Second-Person Standpoint: Respect, Morality and Accountability*, Cambridge (MA): Harvard University Press.

knowledging a certain status, such as addressing the judge in court as "Your Honour". But conflating respect-recognition with behavior leads to conceive respect as forceable, though forced respect sounds contradictory, and certainly does not satisfy the claim to be respected. To my mind, that is the reason why respect for differences cannot be demanded. In conclusion, Balint's work, though prospecting an original reflection on toleration, completely overlooks the symbolic aspect which plays such an important role in the conflicts over diversity and in their proper resolution via principled accommodations.

*University of Eastern Piedmont* ANNA ELISABETTA GALEOTTI

Cameron, Ross P., *The Moving Spotlight. An Essay on Time and Ontology*. Oxford: Oxford University Press, 2015, pp. x + 219.

Time flows. Things change. What is now present was future and will be past. Notoriously, 'A-theorists' (who believe that the passage of time is a genuine feature of reality) have tried to characterise this idea in a rigorous way. One prominent group, the 'presentists', think of the flow of time as a relentless process of creation and annihilation of purely present things. Past and future entities are no part of the inventory of the world.[1] Non-presentist A-theories, by contrast, inflate their ontology with more than merely present things. Some, the 'growing block' theorists, allow for the existence of past things, such as dinosaurs and Roman Emperors. Their inventory of the world becomes bigger and bigger as time goes by, including a growing list of things that were present but are no longer.[2] There is also the "mirror image" of the growing block view, which holds that future things exist, in addition to present ones, but that there are no past things whatsoever. While Caesar is no longer part of the ontological inventory, future Martian outposts are included; the outposts are "out there" waiting to become present. In other words, the flow of time "shrinks" the edge of the block, making the inventory of the world smaller and smaller as time goes by.[3] The last non-presentist A-theory is the 'moving spotlight' view (hereafter, 'MSV'). MSV is a theory according to which 'presentness' is something that *moves*, "somewhat like the spot of light from a policeman's bull's-eye traversing the fronts of the houses in a street. What is illuminated is the present, what has been illuminated is the past, and what has not yet been illuminated is the future".[4] MSV is a version of 'eternalism', the view that past, present, and future

[1] See, e.g., Hinchliff, M. 1996, "The Puzzle of Change", in Tomberlin, J.E. (ed.), *Philosophical Perspectives*, 10, Cambridge (MA): Blackwell, 119-36; Markosian, N. 2004, "A Defense of Presentism", in Zimmerman, D. (ed.), *Oxford Studies in Metaphysics*, 1, Oxford: Oxford University Press, 47-82.

[2] See, e.g., Correia, F. and Rosenkranz, S., 2013, "Living on the Brink, or Welcome Back, Growing Block!", in Bennett, K. and Zimmerman, D. (eds.), *Oxford Studies in Metaphysics*, 8, Oxford: Oxford University Press, 333-50; Forbes, G.A. 2015, "The Growing Block's Past Problems", *Philosophical Studies*, 173, 699-709.

[3] See, e.g., Casati, R. and Torrengo, G. 2011, "The not so Incredible Shrinking Future", *Analysis*, 71, 1-5; Hudson, H. and Wasserman, R. 2009, "Van Inwagen on Time Travel and Changing the Past", in Zimmerman, D. (ed.), *Oxford Studies in Metaphysics*, 5, Oxford: Oxford University Press, 41-49.

[4] Broad, C.D. 1923, *Scientific Thought*, London: Kegan Paul, 59. See, e.g., Skow, B. 2009,

things all exist: Caesar, Lady Gaga, and the Mars outposts all exist, they are all equally real, and they are each located in different parts of the temporal dimension. The view that past and future are equally part of the realm of being is defended also by 'B-theorists'.[5] In contrast to an A-theorist's approach to time, a B-theorist does not take pastness, presentness, and futurity (the 'A qualities') to be part of the fundamental level of reality. No instant can be said to be past, present or future in an absolute sense. Instants of time would be tied together ('ordered') by a mere relation of temporal precedence or succession (the 'B relations'). According to many, the differences between A-theories and B-theories do not prevent philosophers from combining elements of the two approaches. MSV, in particular, is often thought of as exploiting a distinctively B-theoretic ontology (i.e., eternalism) plus the A-theoretic notion of absolute presentness.

Ross Cameron's latest book, *The Moving Spotlight*, takes a step in a different direction. His central, thought-provoking claim is that MSV should be understood as closer to presentism than to a refined version of B-theoretic eternalism. In a nutshell, his idea is that MSV should be conceived as an enriched A-theory, wherein the truth of tensed sentences (e.g., 'Alice is standing' and 'Martha was sitting') rests upon the way things are *now*. And, in accordance with presentism, Cameron's view maintains that there is no difference between how things are and how things are right *now* (162). Nevertheless, Cameron's MSV is genuinely distinct from presentism, since "non-present as well as present entities are some way *now"* (162). In other words,

> the moving spotlighter grants that one can speak from the present perspective *about* the non-present. That one can say how non-present things *now* are. Truth simpliciter is present truth, but amongst the way things are now—*contra* presentism—is that mere past and future entities are some way or other (258).

According to Cameron, this distinctive version of MSV is the best A-theoretic metaphysics on the market. Such a claim might sound puzzling. Famously, and importantly, there are at least *six* problems that a good A-theory should be able to address: (1) the so-called 'epistemological problem' ("How do you know that you are now *now*?"), (2) J.M.E. McTaggart's infamous paradox, (3) a problem of providing adequate truth-makers for past-tensed sentences, (4) a problem of accounting for relations to non-present things, (5) a problem of addressing our intuitions about the openness of the future, and (6) a problem of explaining in what sense the present is 'privileged'. Now, it is widely held that presentism and the growing block view perform better than MSV when dealing with these problems. Growing block theorists are able to deal with (3), (4), (5), and (6), although they struggle with (1) and (2). Presentists, on the other hand, offer an elegant solution to (1), (2), and (6), but face difficulties with the rest. But MSV is usually taken to be in the worst position, overall, since it offers a satisfactory answer to only two of them: (3) and (4).[6] Ross Cameron's aim in *The Moving Spot-*

"Relativity and the Moving Spotlight", *Journal of Philosophy*, 106, 666-78; Skow, B. 2012, "Why Does Time Pass?", *Noûs*, 46, 223-42; Deasy, D. 2015, "The Moving Spotlight Theory", *Philosophical Studies*, 172, 2073-89.

[5] See, e.g., Mellor, D.H. 1998, *Real Time II*, London & New York: Routledge; Oaklander, N.A. 2004, *The Ontology of Time*, Amherst (NY): Prometheus Book; Le Poidevin, R. 2007, *The Images of Time*, Oxford: Oxford University Press.

[6] The fact that presentism and the growing block view solve more problems than MSV is

*light* is to establish that his novel version of MSV successfully addresses all six important problems, in contrast to standard iterations of MSV.

The book is divided into five chapters. In Chapter 1, Cameron argues that, contrary to popular belief, presentism faces the epistemological problem as much as other A-theories. In Chapter 2, he deals with McTaggart's paradox. Cameron concludes that neither the regress nor the circularity identified by McTaggart's argument are vicious; they do not justify the denial of the A-theoretic approach. Chapter 3 explains why MSV is more attractive within an approach to truth-making according to which "giving an ontological underpinning of tense is to say what makes it the case that the tensed truths are true" (24). This approach is opposed to the so-called 'Quine-Lewis-Sider position', according to which "giving an ontological underpinning of tense is to say what it is in tenseless terms for a tensed claim to be true" (23-24). In Chapter 4, Cameron develops (or, at least, tries to develop) a view that, as we anticipated above, shares with presentism the thesis that how things are *now* is how they are *simpliciter*, while inflating the ontology with more than present things. Finally, Chapter 5 analyses how this version of MSV can account for our intuitions concerning the metaphysical difference between a 'fixed' past and an 'open' future.

Is this book worth the read? Yes. Absolutely. At the very least, it is a brilliant defence of MSV. Still, we think there is a crucial point in Cameron's approach that makes his theory obscure, to say the least. As we said above, according to Cameron, to exist simpliciter is to exist now. But, in contrast to the presentist, Cameron accepts that also non-present things are part of reality now, in some way or other. This allows Cameron to defend the claim that his MSV is not a sui generis B-theory, since he does *not* believe in the reality of past or future. What he believes in is the reality of past and future *things*, which can be truly described by saying how they are now, whereas the way they were or will be is not part of reality. Still, Cameron does not seem to offer any account of the way in which past and future things have *now* irreducible past- or future-tensed properties, such as "having been such-and-such". Of course, Cameron might describe the instantiation of a past- or future-tensed property in terms of a *present instantiation* of that same property. But what does it mean exactly? Why should we take an object instantiating a property *now* to be a past or future entity instead of a mere present object? In short, one might have the suspicion that Cameron's view could ultimately collapse into a version of presentism in disguise.[7]

*Centre for Philosophy of Time*
*Department of Philosophy*
*University of Milan*

SAMUELE IAQUINTO
VALERIO BUONOMO

good evidence in favour of the first two theories only under the hypothesis that the problems are equally forceful; and this seems controversial. For example, one might be skeptical on whether, in developing a theory of time, the epistemological concerns raised by the first problem carry the same weight as the metaphysical concerns raised by the remaining ones.

[7] We would like to thank Dave Ingram and Giuliano Torrengo for helpful comments on a previous version of this review.

Dentith, Matthew R.X., *The Philosophy of Conspiracy Theories*.
New York: Palgrave MacMillan, 2014, pp. xiii + 190.

Conspiracy theories are an important, salient aspect of everyone's life, be it vehement rejection of them, passionate advocacy, dispassionate examination, or all three. Be it concerning our personal affairs, corporate activities or political affairs of state. For those who wish to gain the necessary conceptual understanding of the complexities involved in the nature, epistemology and social and political significance of conspiracy theories, philosopher Matthew R.X. Dentith's *The Philosophy of Conspiracy Theories* is a natural starting point. As the title promises, it is, *The Philosophy of Conspiracy Theories*, not *Dentith's Philosophy of Conspiracy Theories*. Yet it seems to me we get the best of both.[1] This is the challenge.

When we take hold of the book, a bright sun-yellow cover, festooned with 37 icons referencing diverse contemporary conspiracy theories greets us.[2] Each recognizable to most anyone, Dentith has perused every one of these families of suspicion. No surprise, the book is rich in real-world examples of conspiracy and conspiracy theorizing, present and past. I was startled one day while walking through a university department to see a poster-sized print of Dentith's book-cover hanging on a colleague's office wall. I should not have been. This colleague is a history professor. Historians know well the reality of ambitious political conspiracies within our collective past.

Dentith has been active in the field of epistemology of conspiracy theory for the last decade. He calls himself a "conspiracy theory theorist". Author of several important papers as well as the blog episto.org and podcast, "The Podcaster's Guide to the Conspiracy", Dentith brings together in this book a thorough research and analysis background. In a discussion as thorough as his, it would be easy, and enjoyable, to write a book about the book.[3]

It opens with a concise forward by New Zealand philosopher Charles Pigden. Pigden's 1995 "Popper Revisited: Or What is Wrong with Conspiracy Theories?" anticipated much of the discussion in the epistemic literature in the years to come.[4] While its prescience was not recognized for a time, now it is considered something of a classic, being the first philosophical literature on the subject in decades. Concerning Dentith's text, Pigden concludes,

> There [is much] to be said on behalf of the great and good who routinely dismiss conspiracy theories as (in Christopher Hitchens' felicitous phrase) the 'exhaust fumes of democracy'. Yes, indeed there is. And that more *is* said—and is said at length—by Matthew Dentith, who patiently refutes it point by point. […] Now, read on (xi).

A forward like this by a major contributor to the literature encourages us to.

[1] While there have been a few fireworks in the field since its publication, all fall easily within the text's categories and explanatory frameworks and appear best understood within these.
[2] Yes, I counted them.
[3] all-embracing.episto.org is a well organized, interesting, even-handed if sometimes intentionally humorous resource for those interested in the epistemology of conspiracy theory. One can also find an imagine of a bright yellow book there.
[4] Pigden, C. 1995, *The Philosophy of the Social Sciences*, 25, 1, 3-34.

Dentith's introduction wastes no time. He immediately points to the "elephant in the room". What say we to conspiracy theories that are, oddly, official government explanations and animate major acts of state and global events? Some conspiracy theories appear ludicrous, some strange but interesting, and some received wisdom. For instance, the official explanation for the 9/11 attacks appears to be a conspiracy theory. After all, people conspired to high jack planes and kill thousands. Yet there is no sense this is suspect for relying on a conspiracy claim. Even if calling an explanation a "conspiracy theory" is, on what Pigden calls "the received wisdom", the kiss of death, a double standard appears to be in play.

We must sort out what is a "conspiracy theory". Then we must determine how we sort out the well evidenced from ludicrous conspiracy theories. This sets the stage for the epistemic adventure ahead. One need not be a specialist in this field of social epistemology—a field that straddles both epistemology and political philosophy—to find the topic every bit as intrinsically interesting as it is socially ubiquitous.

## 1. Definition and Dismissal

Our first meet and greet the elephant brings us to the second chapter, "Conspiracy Theory Theories" and then, like the blind men in the story of exploring the elephant's contours, on to the four chapters that follow. What is the proper definition of "conspiracy theory"? How do we best understand this powerful social practice? Is it mentally defective? Rationally or epistemically deranged? The tensions here seem to derive from the fact that while there are many accusations of conspiracy that are poorly evidenced and implausible, there appear to be a great many conspiracies, both small and extraordinary in their political ambitions, where our belief in them is well evidenced and commonplace. So at what point does belief in a conspiracy become belief in a *conspiracy theory*?

The definitional problem of what a "conspiracy theory" is looms large. It powerfully influences our discussion, as skeptics of "conspiracy theory" will wish to reserve the term to a pejorative, social "kiss of death", while those more epistemic-minded will wish to avoid *a prior* mal-biasing against explanations that cite a conspiracy as a significant cause of events, recognizing that in many cases such explanations turned out to be correct. Much has been written on the definitional issue in recent years, a fight largely over what the term "conspiracy theory" is a gateway to; delusion *or* intellectual honesty? Definitional issues may seem uninviting, but in Dentith's hands the survey he supplies proves to be an excellent teaser for anyone interested in the nature of different social explanations.

On the question of definition of "conspiracy theory", Dentith's solution is elegant: A conspiracy theory is any explanation that cites a conspiracy as a salient cause. Its inclusive nature opens up the great canyon-lands concerning our relationship to conspiracy and its theory as something ubiquitous. Epistemic and pragmatic questions follow about the proper strategies to distinguish which conspiracy theories to investigate. These questions would have been invisible under more constrained, politically distorted approaches to the issue. This is just what we would hope from good philosophical work; answers that give us better questions.

Next, Dentith outlines the basic fork in the adventure. The first tine: Is there something *mentally* or *socially* misguided about belief in conspiracy explanations? Pejorative definitions of conspiracy explanations portray them as somehow fundamentally flawed, and those who explore them, pathological; pejorative, pathologizing attitudes do the same. Consider the claim by social psychologists Brotherton and French that a conspiracy theory can be defined as *an unverified and relatively implausible allegation of conspiracy*. They use a variety of different definitions to try and capture what they are referring to with respect to conspiracy theories. Aside from the claim conspiracy theories are unverified and relatively implausible, they also classify them as "anomalistic beliefs", which they define as "[beliefs] that defy conventional understanding of reality, including (but not limited to) belief in the paranormal and conspiracy theories".[5]

The tactic of pejorative definitions and attitudes is to divert the discussion from epistemic issues to psychological critiques and sociological fears. Set against the background of our epistemic and political concerns, this diversion emerges as interesting but comparatively limited in importance. It is, however, easily abused. Dentith poses insightful and pressing questions for us to consider about any psychological, pathologizing, maneuver. He is not alone in this concern about how to frame our understanding of conspiracy theorizing. An establishmentarian assumption seems at work in pejorative glosses on "conspiracy theory"; one at odds with human history and normal human rationality. So Dentith's discussion is useful for social scientists not just interested in how to approach conspiracy *theories*, but conspiracy *theorists*.

To the second tine: Is there something intrinsically *epistemically* defective in conspiracy theories? Again, the discussion is interesting and informed. How we easily separate warranted conspiratorial explanations from warranted non-conspiratorial merely by pointing to the explanatory structure of either. The puzzle, if there is one, is that as explanations based on human intentions and subsequent actions, both appear the same. But are they?

## 2. Social Epistemology

All the proceeding leads us to the critical epistemic discussion. Beginning with chapter 6, and for six more chapters, the book turns to a detailed exploration of the epistemology of how and when we should, or should not, embrace a conspiracy theory. This discussion spans almost ninety pages of carefully delineated material and is an excellent introduction to the fascinating debates here: The crux of the epistemology of conspiracy theories and conspiracy theorizing, really. I think it is safe to say: If we want to understand humans—a social and highly organized, hierarchical, cooperative and deceptive primate—we have to understand our practice of conspiracy. That inevitably makes us conspiracy theorists. Even theorists of conspiracy theory.

We operate on both the particular and the epistemic level. Establishing consistency here has proven difficult for many; especially when it troubles our current political pieties. Dentith's approaches to this include the problems with appeal-to-authority arguments against conspiracy theories, whether official stories have any privileged epistemic status (I agree, they often do not), and how

[5] Brotherton, R. and French, C. 2014, "Belief in Conspiracy Theories and Susceptibility to the Conjunction Fallacy", *Applied Cognitive Psychology*, 28, 2, 238-48 (239).

evidence and inference functions in the evaluation of conspiracy theories. In short, a fore-taste of social epistemic heaven. Whatever we make of the critiques deployed by Dentith and the conclusions reached, the discussion is wide-ranging, informative and sometimes daring.

Once the way has been made for the rational legitimacy of conspiracy theories, we need to know how to judge them epistemically. The answer that emerges is: *Only on their evidential merits*. In the literature it has come to be known as "particularism" about conspiracy theories. They should be judged on the evidence particular to each. Two young philosophers, Taylor and Buenting, coined this term for "case by case" evaluation in 2010.[6] But the currency of this simple moniker is owed to Dentith. In sum, Dentith's view is that conspiracy theories turn out to be birds ordinary to the flock of social explanations, and there is no *general* reason to be skeptical about them. Particularism, while it may seem obvious enough, is revolutionary when placed in contrast to the long winter of automatic, if irresponsible, dismissal of conspiracy theories that characterized most of 20th century academia and mainstream social commentary.

When turning to the epistemic issues, the book shines brightest, rather like its cover. Dentith supplies his readers a God's eye view to the epistemic debate about conspiracy theories. The recent research discussed includes that of Charles Pigden, Brian Keeley, David Coady, Steve Clarke, myself and others. This research defines the field as we find it today. He covers the work of these social epistemologists with scrupulous attention to detail and an unswerving fidelity to their actual positions. Dentith's critiques in response to the diverse positions within this wide-ranging debate are original and important. The more one is familiar with this debate, the more one recognizes this. His responses should be of interest to philosophers wishing to join what proves to be an exciting and socially relevant discussion.

If we have some reservations about Dentith's approach to epistemic impasses, these are often evidence of his originality and caution. For instance, *attenuation strategies* are frequent in Dentith's work by "taking the next step" and arguing while the problems are real, they are not as bad as they might seem. This is the opposite of my typical concern, which is to directly critique the current information hierarchy's basic methods of information distribution. We ought to regard these as sometimes unreliable, and more likely to be unreliable when we most need them to be reliable. Dentith does a fairly good job of conceding the difficult epistemic problems we face, but is careful to not allow the spectacle to overwhelm. Or naiveté to seduce us. This is unusual in the literature. It is an interesting manoeuvre. Instances of this "taking the next step" in order to attenuate are found throughout the book.

Juha Räikkä has maintained that our use of "conspiracy theory" is dependent on its contrast to the established official narrative. Räikkä suggests conspiracy theories lose their property of being such when they gain sufficient acknowledgement, popular or at least, official. No contrast, no conspiracy theory.[7] David Coady takes a similar if more basic approach, requiring that a conspiracy

[6] Buenting, J. and Taylor, J. 2010, "Conspiracy Theories and Fortuitous Data", *Philosophy of the Social Sciences,* 40, 4, 576-78.

[7] Räikkä, J. 2009, "The Ethics of Conspiracy Theorizing", *Journal of Value Inquiry*, 43, 457-468, 460.

theory by definition must be contrary to an official story (2006).[8] There are numerous counter-examples to this claim—for instance, when a government has no official story concerning x, yet people are promoting conspiracy theories concerning x. Both Coady and Räikkä's claims are at best statistical ones about linguistic usage, and therefore malleable and amenable to correction. Dentith comments,

> However, we should ask why we—particularly philosophers—would want to preserve common usage if it does not advance our analysis of these things called 'conspiracy theories'. We can add to this that Coady's defense of this particular common usage might also have the negative effect of enabling government conspiracies. If we preserve the notion that the terms 'conspiracy theories' and 'conspiracy theorists' are pejoratives, then that might shield conspirators from the accusation that they are conspiring (113).

Dentith gives ground, attenuating the problem, agreeing that "conspiracy theory" is understood as "contrary to the official narrative" but asks that (1) we change this understanding in the name of better analysis and (2) we presently ignore this purported common usage as it is dangerous in a democracy.

My approach to Coady's "contrary" addendum is direct. I suspect "contrary to official stories" is not common usage. The claim that it is contrary to common usage is contrary to my experience, except in certain academic parlors and within mainstream media. Research by social psychologist Michael Wood supports this. There is no popular correlation between "conspiracy theory" and "unlikely".[9] In Wood's study participants were presented the same official narrative, but in one version, "conspiracy theory" was the representation of political facts, in the other it was not. No differences in credibility occurred. We might conclude the pathologizing nature of the phrase "conspiracy theory" is an interesting figment of mainstream media, political orthodoxy and academia. By "official stories" we mean the accounts of mainstream media, government and orthodox academia. As a common user I doubt that in common usage "conspiracy theory" is necessarily contrary to official stories, or even ordinarily contrary to them. Further, factions within government, corporations or families can talk about other factions conspiring against them and hardly violate common usage. Others may then denounce these claims as "conspiracy theories"; but there is no official story to be contrary to.[10] Yet, popularly, we tend to discuss those that are

[8] Coady, D. 2006, "Conspiracy Theories and Official Stories", in Coady, D. (ed.), *Conspiracy Theories: The Philosophical Debate*, Hampshire: Ashgate, 115-27 (117).

[9] See Wood, M.J., "Some Dare Call It Conspiracy: Labeling Something a Conspiracy Theory Does Not Reduce Belief in It": http://onlinelibrary.wiley.com/doi/10.1111/pops.12285/full. The upshot is "conspiracy theory" does not function as a term of popular dismissal or contrary to an "official story".

[10] In my "Conspiracy Theory and Rationality", in Jensen, C. and Harre, R. (eds.), *Beyond Rationality*, Newcastle upon Tyne: Cambridge Scholars Publishing, 2011, I offer a more detailed counter-example to Coady's addendum involving a series of murders which someone comes to recognize as connected to a conspiracy, and thus forms a conspiracy theory; but there is no official story, just that this murder occurred, that murder occurred, and so on, and the conspiracy theory she forms is hardly contrary to *that*. It is consistent with that and relies on it. There is nothing that violates common usage here, so the best explanation appears to be that Coady is mistaking meaning, or if you like, common usage, with salience.

contrary to official accounts as these are more *salient* to us. Coady and Dentith miss this. Salience is what appears to be at work, not common usage or meaning. Coady is not respecting common usage; he is paying homage to a political doctrine about discourse in public venues: Do not violate the official narrative of Western democratic society. The story of the emperor's new clothes comes to mind. However, Dentith's attenuation technique is very useful in academia. The opposite of tone-deaf. In academia an overt anxiety exists concerning conspiracy theory, as it under-cuts the tacit assumption—a rather strange and suspect one, given well know history, distant and recent—that the established political and economic order is more or less proper, only conspiratorial when malfunctioning (or protecting the state against other states) and our proper analysis of it should never question this, but always presuppose it.

## 3. Conclusion

Anyone interested in the questions surrounding conspiracy theory and theorists will find Dentith's book a refreshingly clear, calm and thorough accounting and analysis of the issues concerning conspiracy theory and how we can approach these. It is an adventure. Again, the sign of an interesting book, one about an important social topic at a high level of analysis and honest manner of examination, is the way it provokes diverse questions and disagreements. Dentith's even-handed style while playing with fire does not disappoint on this score. His intellectual caution will appeal to those who are new to the field. His forthrightness will appeal to all of us. When you have gained far better questions and lost simplistic answers, Dentith knows he has done his job.

*South Texas College* LEE BASHAM
*University of Texas*
*Rio Grande Valley*

# Advisory Board