Philosophers' Imprint

volume 18, no. 15
august 2018

Reflection Principles and the Liar in Context

Julien Murzi and Lorenzo Rossi

Philosophy Department (KGW), University of Salzburg, and Munich Centre for Mathematical Philosophy, Ludwig-Maximilians University
Philosophy Department (KGW), University of Salzburg

Abstract

Contextualist approaches to the Liar Paradox postulate the occurrence of a context shift in the course of the Liar reasoning. In particular, according to the contextualist proposal advanced by Charles Parsons (1974) and Michael Glanzberg (2001, 2004a), the Liar sentence λ (asserting that λ does not express a true proposition) doesn't express a true proposition in the initial context of reasoning c, but expresses a true one in a new, richer context c', where more propositions are available for expression. On the further assumption that Liar sentences involve propositional quantifiers whose domains may vary with context, the Liar reasoning is blocked. But *why* should context shift? We argue that the paradox involves principles of *contextualist reflection* that explain, by analogy with well-known reflection principles for arithmetic, why context must shift from c to c' in the course of the Liar reasoning. This provides a *diagnosis* of the Liar Paradox—one that equally applies to two *revenge arguments* against contextualist approaches, one recently advanced by Andrew Bacon (2015), the other mentioned by Charles Parsons (1974) and more recently revived by Cory Juhl (1997).

Keywords: Liar Paradox · Contextualism · Reflection principles · Revenge Paradoxes · Absolute generality

Contextualist approaches to the Liar Paradox seek to preserve classical logic and a version of the naïve principles governing the use of 'true'. They do so by postulating the occurrence of a context shift in the course of the Liar reasoning.[1]

In particular, according to the contextualist proposal first advanced

1. Let λ be a sentence identical to 'λ is not true'. We can then reason that if λ is true, then λ holds, whence λ is not true. However, if λ is not true, then λ again holds, whence λ is true. That is, λ is true if and only if it isn't—a contradiction in most logics. The existence of sentences such as λ can be proved given minimal syntactic assumptions, and hence cannot be plausibly questioned (see e.g. Heck, Jr. (2007, 2012, 2013)). As for the Liar reasoning, it only relies on basic classically valid principles and on the interderivability of φ and 'φ is true'. But how can one plausibly relinquish either of these ingredients? This is the Liar Paradox.

by Charles Parsons (1974) and more recently defended by Michael Glanzberg (2001, 2004a, 2015), the Liar sentence λ should be understood as the the claim that λ doesn't express a true proposition. Then, these theorists maintain, λ doesn't express a true proposition in the initial context of reasoning c, but expresses a true one in a new, richer context c', where more propositions are available for expression. On the further assumption that λ, thus understood, involves propositional quantifiers whose domains may vary with context, the Liar reasoning is blocked: even if λ is untrue in c, it can be consistently evaluated as true in c'. But *why* should context shift from c to c'? According to Christopher Gauker (2006), contextualists cannot satisfactorily answer this question and, for this reason, contextualist approaches to the Liar are hopelessly *ad hoc*. As he puts it, contextualists suppose that 'the context in which we judge that the liar sentence is true is not the context in which the liar sentence says of itself that it is not true' but 'we have no good reason to think that we can 'step back' and judge the liar sentence to be in some way true' (Gauker, 2006, p. 393).[2]

We don't share Gauker's pessimism. We criticise Glanzberg's own account of context shift, according to which a specific use of the predicate 'expresses' triggers a shift in context. We then argue that the paradox involves contextualist reflection principles that explain, by analogy with well-known reflection principles for arithmetic, why the context must shift from c to c' in the course of the Liar reasoning. We suggest that, in Charles Chihara's terminology, this supplies a diagnosis of the Liar Paradox (Chihara, 1979, p. 590). To anticipate a little, the paradox effectively involves the use of a reflection principle that clashes with the lesson standardly drawn from Löb's Theorem. At the same time, this suggests a natural solution to the paradox—one according to which the use of a reflection-like principle leads to a deductively stronger context of reasoning, and hence shifts context.[3]

It is sometimes alleged that any attempt to solve the semantic

2. For a similar line of criticism, see also Rumfitt (2014, pp. 46-7).

3. We should note at the outset that we claim no originality in suggesting a connection between contextualism and reflection principles. For instance, Parsons (1974, pp. 384) briefly hints at reflection principles when discussing sentential truth-theoretical principles such as $\mathsf{Tr}(\ulcorner \varphi \urcorner) \to \varphi$, and Glanzberg (2004b) appeals to reflection principles in arithmetic in an argument aimed at opposing the objection that hierarchical theories fragment the notion of truth. However, neither Parsons nor Glanzberg appears to think that reflection principles for formal arithmetic can help explaining the occurrence of a context shift in the Liar reasoning. Reflection of the kind exhibited by reflection principles for arithmetic must also be distinguished from the kind of semantic reflection extensively discussed in Glanzberg (2015). Whereas reflection in our sense involves a theory's capability to assert its own soundness, Glanzberg's notion of semantic reflection is rather the process of making explicit the complex relations of reference at work in a Tarskian definition of truth. For a comprehensive survey on reflection principles, see Beklemishev (2005).

paradoxes faces unescapable revenge arguments: arguments aimed at showing that any consistent (or at least non-trivial) theory of truth and other semantic properties lacks the resources to block, let alone diagnose, paradoxes that are similar to, but distinct from, the original Liar.[4] We consider two possible revenge arguments against contextualist theories: one recently advanced by Andrew Bacon (2015), the other mentioned by Charles Parsons (1974) and more recently revived by Cory Juhl (1997). We suggest that our diagnosis extends to these alleged revenge paradoxes and precisely points to where both these arguments go wrong.

The discussion is structured thus. §1 introduces the basic features of contextualist theories of truth. §2 presents the Liar in context. §3 criticises Glanzberg's salience-based account of the context shift in the Liar reasoning. §4 motivates our alternative explanation involving a contextualist form of reflection. §5 focuses on revenge. §6 concludes.

1. Contextualist approaches to truth and paradox

We begin by introducing the main ingredients of the contextualist treatment of the Liar Paradox found in Parsons (1974) and Glanzberg (2001, 2004a). In doing so, we fill in a number of details that are left unspecified in Parsons' and Glanzberg's presentations, thus oftentimes departing from Parsons' and Glanzberg's letter.

4. See e.g. Scharp (2013, Chapter 4).

The basic contextualist idea is this. Certain sentences *say different things*, or *express different propositions*, in different contexts.[5] This phenomenon, context dependence, occurs whenever a sentence contains expressions whose value varies with context.[6] The canonical list includes expressions such as 'I', 'now', 'here', and so on. A slightly less canonical, but still standard, list also includes the quantifiers 'every' and 'some'. This is because, arguably, domains of quantification are relative to context.[7] Consider, for instance, sentences such as

(1) There's nothing left in the fridge.
(2) Everybody went to the party.

On their intended interpretations, the quantifiers in (1) and (2) are both restricted to contextually determined domains of quantification: things that can be eaten, certain people, and so on. According to the contextualists, the Liar sentence is context dependent in the same way as (1) and (2).[8] For one thing, according to the contextualist theories we will focus on in this paper, λ is to be understood as a sentence identical to 'there is no true proposition expressed by λ' and, we are assuming, quantifiers are context dependent.[9] For another, these theories postulate the occurrence of a *context shift* in the course of the Liar derivation—one that is determined by a change in the domain of quantification of the propositional quantifiers.[10]

We assume that propositions are the primary bearers of truth and falsity. Accordingly, a sentence φ is true in a context c if and only if it expresses a true proposition in c (Kaplan, 1989, p. 522). We formalise the right-hand side of this biconditional as follows:

$$\exists_c p(Exp(\ulcorner\varphi\urcorner, p, c) \wedge \mathbf{Tr}(p)),$$

where $Exp(\ulcorner\varphi\urcorner, p, c)$ reads 'φ expresses p in c', $\exists_c p$ expresses existential quantification over a domain of propositions determined by c, **Tr** expresses propositional truth, and $\ulcorner\varphi\urcorner$ is a name of φ (for instance, $\ulcorner\varphi\urcorner$ can be taken to be the Gödel code of φ in some standard Gödel coding). For simplicity's sake, we keep our presentation largely informal and do not introduce a formal language for contextualist theories. We call the index c appearing in $\exists_c$ and $\forall_c$ and the term c appearing in expressions of the form $Exp(\ulcorner\varphi\urcorner, p, c)$ *context parameters*.

Throughout the paper, we assume the following intended reading of the context-sensitive elements of the language, namely the quantifiers $\exists_c$ and $\forall_c$ and the relation *Exp*. Recall, a sentence might express a proposition in a given context c and another proposition (or no proposition at all) in another context c'. Accordingly, the role of context parameters is to indicate which propositions are available for expression in a given context. We say that a context c' *extends* a context c if the propositions available for expressions in c are also available for expressions in c'. Moreover, a context c' *strictly extends* c if it extends c but c does not extend c'. It follows from this definition that the extension relation is a partial order, i.e. reflexive, transitive, and antisymmetric.[11] In this paper, we do not provide a fully worked out semantics for the contextualist language; we only provide schemata governing the use of the truth and expressibility predicates. However, it will be useful to bear in mind the intended interpretation of the constitutive elements of the language we have just sketched

5. We talk of propositions purely for the sake of convenience. Nothing in our argument hangs on their existence or properties: our arguments can all be recast in terms of interpreted languages (Parsons, 1974).
6. See e.g. Kaplan (1989).
7. For some recent accounts, see e.g. Stanley and Szabo (2000) and Gauker (2010).
8. See e.g. Parsons (1974); Burge (1979); Glanzberg (2001, 2004a).
9. Some theorists locate the context-sensitivity in the truth predicate itself (e.g. Burge, 1979). However, we don't find this approach convincing, since there do not seem to be independent reasons to accept that 'true' is context dependent.
10. We should note that, in Glanzberg's view (Glanzberg, 2004a, p. 30 and ff.), the quantifiers 'every' and 'some' exhibit an 'extraordinary' kind of context dependence. However, Glanzberg's distinction between different kinds of context dependence is not relevant for the purposes of the present paper.
11. This is all we assume about the general structure of contexts. In particular, we do not require that for any two given contexts c and c' one strictly extends the other, nor that we can always find a context c^* that strictly extends all the contexts in an arbitrary collection.

We interpret contextualists as adopting a collection of context-relative theories $S_c, S_{c'}, S_{c''}$, and so on, whose derivability relations are indicated as follows:

$$\vdash^c_S, \vdash^{c'}_S, \vdash^{c''}_S, \ldots$$

In order to introduce and derive the Liar in context, we take each contextualist theory S_{c^+} to be the smallest set of sentences such that:

- $S_c \subseteq S_{c^+}$, for every c extended by c^+.
- It contains all the instances of the following schemata:

(CTS) $\vdash^c_S \forall_c p[Exp(\ulcorner\varphi\urcorner, p, c) \rightarrow (\varphi \leftrightarrow \mathbf{Tr}(p))]$
(UN) $\vdash^c_S \forall_c p \forall_c q((Exp(\ulcorner\varphi\urcorner, p, c) \wedge Exp(\ulcorner\varphi\urcorner, q, c)) \rightarrow p = q)$

for c extended by c^+.
- It is closed under the following rule:

(NEC) If $\vdash^c_S \varphi$, then $\vdash^c_S \exists_c p(Exp(\ulcorner\varphi\urcorner, p, c) \wedge \mathbf{Tr}(p))$

for c extended by c^+.
- It is closed under classical logic.

Let us now examine these principles more closely.[12] CTS is a contextually restricted, and yet intuitive, version of the T-Schema for sentential truth:

(TS) $\varphi \leftrightarrow \mathsf{Tr}(\ulcorner\varphi\urcorner)$,

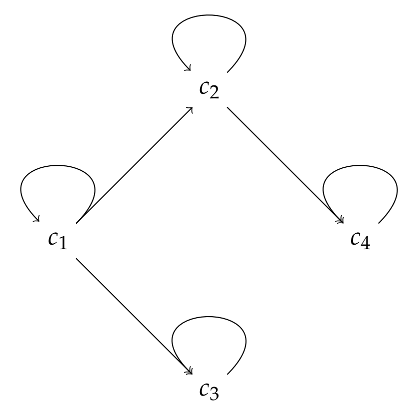
where Tr is a sentential truth predicate. It states that if φ expresses a proposition p in some context c, then p is true if and only if φ. The principle UN, for *uniqueness*, ensures that there is at most one proposition expressed by a sentence φ in a context c. Finally, NEC establishes that if S_c proves φ, then S_c proves that φ expresses a true proposition in c.

It is immediate to see that each contextualist theory S_c satisfies the following requirements:

(C-DER) If $\vdash^c_S \varphi$, then c extends the context parameters in φ;
(EXT) If $\vdash^c_S \varphi$, then $\vdash^{c'}_S \varphi$, if c' extends c.

Both of these principles will play an important role in the arguments to be given in §4.

It follows from the definition of the contextualist theories S_c that the index c in S_c and $\vdash^c_S \varphi$ is a bookkeeping device that keeps track of the instances of the contextualist principles that are in S_c. That is, each S_c only contains instances of CTS and UN, and is closed under applications of NEC to sentences, whose context parameters are extended by c. For example, consider four contexts, c_1, c_2, c_3, and c_4 ordered by extensibility as shown in the following graph:



where an arrow from c_i to c_j indicates that c_j extends c_i. In this example, S_{c_4} contains all the instances of CTS and UN with context parameters c_1, c_2, and c_4, while S_{c_3} only contains all the instances of CTS and UN with context parameters c_1 and c_3 (similarly for the applications of NEC).

It might be thought that context plays a double role—one syntactic, the other semantic. On one hand, c is the bookkeeping device that we

12. See Parsons (1974, p. 387) and Glanzberg (2001, §2).

have just described. On the other, it could also be read as indicating a collection of propositions available for expression. However, we should stress that our main concern in this paper is with syntactic principles governing truth and expression in context. That is, we do not give a proper semantics for the theories we discuss—we only offer, for the sake of clarity, an informal interpretation of the context-sensitive elements of the language. We now turn to the contextualist treatment of the Liar Paradox.

2. The Liar in context

Let k be the initial context of reasoning for the following derivation. The paradox involves a sentence λ equivalent to the sentence 'λ doesn't express a true proposition in k':

$$\lambda \leftrightarrow \neg\exists_k p(Exp(\ulcorner\lambda\urcorner, p, k) \wedge \mathbf{Tr}(p)).$$

The Liar in context can then be presented thus (see Glanzberg, 2004a, pp. 33-4):[13]

(3) Suppose (in k) that λ expresses a proposition in k—call this proposition q.

(4) Suppose $\mathbf{Tr}(q)$.

(5) Then, λ. [3, 4, CTS]

(6) So there is no true proposition expressed by λ in k. [Definition of λ]

(7) Hence, $\neg\mathbf{Tr}(q)$. [4-6, UN]

(8) Suppose $\neg\mathbf{Tr}(q)$.

(9) Then, $\neg\lambda$. [3, 8, CTS]

(10) So there is a true proposition expressed by λ in k. [Definition of λ]

(11) Then, $\mathbf{Tr}(q)$. [8-10, UN]

(12) $\mathbf{Tr}(q)$ if and only if $\neg\mathbf{Tr}(q)$ [4-7, 8-11]

(13) Therefore, λ does not express a proposition in k. [3-12, logic]

(14) But, then, it does not express a *true* proposition in k. [13, logic]

(15) Then, λ. [Definition of λ]

(16) Thus, λ expresses a true proposition in k. [15, NEC]

(17) Contradiction. [14, 16]

Contextualists maintain that the foregoing argument is invalid.[14] In their view, a *context shift* takes place between (13) and (16), so that (16) is:

(16*) λ expresses a true proposition *in* k' (for a context k' that strictly extends k).

Since (13) and (16*) are perfectly consistent, the Liar reasoning is blocked.

Before turning to the question why, according to contextualists, context shifts in the course of the Liar derivation, it is important to notice that contextualists take the derivation (1)-(16*) at face value. In their view, λ expresses a proposition in k', but not in k. That is, the domain of propositions on which the propositional quantifier $\exists_c p$ ranges must be seen as *expanding*. Where W and W' are the collections of propositions available for expression in, respectively, k and k', what is required is that there be, in W', 'a proposition for λ to express, that could not be any subset of W. This requires [...] an expanded W' available [in k']' (Glanzberg, 2001, p. 239; Glanzberg's notation has been adapted to

13. Our version of the Liar in context is slightly different from Glanzberg's. Glanzberg resorts to a weaker version of NEC, according to which if φ is derived in c, then φ expresses a proposition in c. Having established that λ does not express a proposition in k, and hence, *a fortiori*, a true proposition, Glanzberg concludes via the weaker version of NEC that λ expresses a proposition in k after all. Our version of the Liar in context is more in line with standard non-contextualist presentations of the Liar Paradox, in which λ is identical to $\neg\mathsf{Tr}(\ulcorner\lambda\urcorner)$ and both $\mathsf{Tr}(\ulcorner\lambda\urcorner)$ and $\neg\mathsf{Tr}(\ulcorner\lambda\urcorner)$ are proved using both directions of the sentential T-Schema.

14. See e.g. Parsons (1974); Glanzberg (2001, 2004a).

ours). The reasoning generalises: as Glanzberg puts it, the contextualist approach effectively presupposes 'an *open ended* hierarchy of contexts, and [propositions] available in them' (Glanzberg, 2001, p. 240). We return to this key point in §3, at the end of §4, and in §5.2 below. For the time being, we turn to two key questions: *where* does context exactly shift, and *why* should it shift in the first place?

3. Salience

In his first article on the Liar in context, Glanzberg suggests that context shifts between (13) and (14), which he respectively labels (A) and (B). He writes:

> Let us consider [. . .] the problematic inference. The crucial steps are:
>
> (A) The conclusion that an utterance of λ does not express a proposition.
>
> (B) From (A), the conclusion that an utterance of λ does not express a true proposition.
>
> At (B), we assert λ and express a proposition, and so run into paradox . . . to avoid the Paradox, we must recognize a context shift between (A) and (B). (Glanzberg, 2001, p. 233)

In his second article on the subject, Glanzberg locates the context shift between (13) and (15), i.e. between the derivation of $\neg\exists_k p(Exp(\ulcorner\lambda\urcorner, p, k))$ and the derivation of λ. Glanzberg writes that because both (13) and (15) are

> the results of sound proofs, they must both be true. But the truth of the first requires that there be no proposition for λ to express, while the truth of the second requires that there be one. Hence the paradox. (Glanzberg, 2004a, p. 34)

In Glanzberg's view, both (13) and (15) are true, and hence consistent. But, Glanzberg argues, this can only be so if context shifts between (13) and (15).

However, it is hard to see where context could plausibly shift between (13) and (15). After all, (14) follows from (13) by pure logic and $\neg\exists_k p(Exp(\ulcorner\lambda\urcorner, p, k) \wedge \mathbf{Tr}(p))$ is equivalent to λ. But logical rules and the syntactic principles guaranteeing the equivalence of λ and $\neg\exists_k p(Exp(\ulcorner\lambda\urcorner, p, k) \wedge \mathbf{Tr}(p))$ can hardly have context-shifting properties. What is more, it is difficult to see how context could plausibly *expand* between (13) and (15). Line 13 requires that there be no proposition expressed by λ in k. Since (13) has been proved, we therefore conclude that there are no propositions expressed by λ in k. But then, the truth of (14) is also grounded on the fact that there are no propositions expressed by λ in k, and so is that of (15), since by construction of λ the contents of (14) and (15) are identical. *Pace* Glanzberg, the consistency of either (13) or (14), on one hand, and (15), on the other, doesn't require a shift of context—let alone one in which the Liar sentence can be seen to express a proposition after all.

But let us see why, according to Glanzberg, context shifts between (13) and (15). Glanzberg assumes, plausibly enough, that 'context provides a running record of what is salient; particularly, what is salient in a discourse at a particular point' (Glanzberg, 2004a, p. 37). He then argues that context shifts between (13) and (15) because of a change in the salience structure of the context. His argument is that because (13) is 'the first point in the proof above where there are no undischarged premises' this is the point where

> the [expressibility] relation Exp (which interprets *Exp*) is accepted as salient in the discourse. But then, with the assertion (A), the salience structure is expanded to include Exp. The context thus shifts, by expanding its salience structure, so we have a genuine difference in context between (A) and (B), just as the paradox requires. (Glanzberg, 2004a, pp. 39-40)

However, if (13) is the first occurrence of *Exp* from no undischarged premises, and this induces a change in the salience structure of the context, then *context should already shift there*. To see why, consider the following sentences:

(18) I am cold

(19) I am not cold

Suppose (18) and (19) are respectively uttered by Lisa and Michael. Then, in (18)'s context of utterance, the indexical 'I' designates the speaker of the context, viz. Lisa, and in (19)'s context of utterance, the indexical 'I' designates the speaker of the context, viz. Michael. Crucially, it is Michael's utterance of (19) that changes the context. By uttering (19), Michael becomes the speaker of the context: hence, the context changes, and 'I' is interpreted as designating him, rather than Lisa. Similarly for the Liar in context. If, as Glanzberg suggests, the first undischarged occurrence of *Exp* at line 13 changes the salience structure of the context, thereby inducing a change of context, context must shift at line 13, just like the first occurrence of 'I' as uttered by a new speaker in (19) shifts context at line 19.[15]

But can context shift at line 13? If the context shift takes place at the first undischarged occurrence of *Exp*, then (13) is inferred in k'. But then, line 13 should really be

(13*) λ does not express a proposition in k'.

However, the difficulty now is that (13*) cannot be correctly inferred via the rule of negation introduction from a derivation of a contradiction from (3), for the simple reason that (13*) is *not* the negation of (3). That is, if in the Liar reasoning context shifts at the first undischarged occurrence of the predicate *Exp*, we are forced to conclude that *context doesn't shift*, since the sentence in which such a context-shifting occurrence is meant to occur cannot be correctly inferred from the subproof (3)-(12).

15. We should stress that we don't take our example involving (18) and (19) to be a model of the kind of context shift invoked by Glanzberg. Glanzberg explains the expansion of salient structure by analogy with the introduction of new salient terms in the discourse. His examples involve anaphoric constructions that enable one to fix a referent, as in 'I broke a wine glass last night. It was expensive' (Glanzberg, 2004a, p. 37). However, the argument just given doesn't require an example involving anaphoric constructions.

It might be thought that context shifts at line 14 instead. But there are at least two problems with this suggestion. First, if the argument given two paragraphs back is correct, we lack a reason for thinking that context shifts there. Second, the occurrence of a context shift at (14) would square badly with contextualist orthodoxy. To see this, let us concede that, implausibly, context shifts at line 14, i.e. *after* the first occurrence of *Exp* from no undischarged premises. Then, (13) effectively entails

(14*) λ does not express a true proposition in k'.

But this is as far as the Liar derivation goes. On our assumptions, the Liar sentence λ can no longer be inferred from line 14* of the paradoxical derivation, since λ is identical to line 14, *but not to line* 14*. And while this move blocks the derivation of a contradiction via the Liar reasoning, it sits very poorly with contextualist wisdom. According to contextualists, we learn from the Liar derivation that λ does not express a proposition in the initial context of reasoning and later expresses a proposition in a new, richer context.[16] Recall, this was the contextualist's main motivation for thinking that contexts and domains of propositions form an open ended hierarchy. However, the view that (14*) as opposed to (14) follows from (13) undercuts the contextualist's motivation for thinking this. If (14*) follows from (13), all that follows from the Liar reasoning is that λ doesn't express a true proposition in some context k' different from the initial context of reasoning k. We conclude that Glanzberg's account of context shift is inadequate. It explains neither where nor why context shifts in the course of the Liar reasoning.

4. Reflection

If context cannot plausibly shift between (13) and (15), as we have argued in §3, and if it must shift between (13) and (16), as contextualists maintain, there is only one option left: context must shift between (15)

16. See e.g. Glanzberg (2004a, pp. 34-35).

and (16). That is, it is the passage from the proof of φ in c to the proof that φ expresses a true proposition in c that shifts context in the course of the Liar derivation. This requires that NEC be rejected and replaced by the following principle:

(C-NEC) If $\vdash^{c}_{S} \varphi$, then $\vdash^{c'}_{S} \exists_{c'} p(Exp(\ulcorner\varphi\urcorner, p, c') \wedge \mathbf{Tr}(p))$,

for some context c' that strictly extends c.[17,18] From this section on, we therefore narrow down our focus to contextualist theories formulated as explained in §1, but with closure under NEC replaced by the constraint that if φ is proved in S_c, then it can be proved in $S_{c'}$ that it expresses a true proposition in c', where c' extends c.

But why should NEC be rejected, apart from the fact that it is one of the (many) ingredients of the Liar in context? What, if anything, is wrong with this principle? And why should C-NEC be at all plausible? In particular, why should the fact that φ is proved in c entail that it expresses a true proposition in some *other* context? On the face of it, the principle doesn't sound very intuitive. As Gauker puts it:

> Might we nonetheless 'step back' and assert that the sentence in (14) expresses a [true] proposition in some *other* context? [. . .] May we conclude that λ expresses a proposition in k'? Perhaps, but we lack a reason to think so. (Gauker, 2006, p. 403, original notation adapted to ours)

If context shifts between (15) and (16), NEC must be replaced with C-NEC. But the latter principle would appear to be *ad hoc*: it postulates the occurrence of a context shift, but cannot explain why such a shift should take place. Gauker concludes that context doesn't after all shift in the course of the Liar reasoning. We aim to show otherwise. More precisely, we suggest that C-NEC can be justified via an analogy with *reflection principles* in arithmetic (the reader familiar which such principles can skip the next two paragraphs).

Let T be a recursively axiomatisable theory and let $\mathsf{Bew}_T(\ulcorner\varphi\urcorner)$ express the existence of a proof of φ from the axioms of T using the rules of T, where $\ulcorner\varphi\urcorner$ is a code of φ relative to some suitably chosen coding scheme.[19] In order for Bew_T to be a standard provability predicate, it must obey the following conditions, known as Löb's derivability conditions:

(L1) If $\vdash_T \varphi$, then $\vdash_T \mathsf{Bew}_T(\ulcorner\varphi\urcorner)$.

(L2) $\vdash_T \mathsf{Bew}_T(\ulcorner\varphi \to \psi\urcorner) \to (\mathsf{Bew}_T(\ulcorner\varphi\urcorner) \to \mathsf{Bew}_T(\ulcorner\psi\urcorner))$.

(L3) $\vdash_T \mathsf{Bew}_T(\ulcorner\varphi\urcorner) \to \mathsf{Bew}_T(\ulcorner\mathsf{Bew}_T(\ulcorner\varphi\urcorner)\urcorner)$.

Perhaps the most well-known reflection principle is the *local reflection principle*:

(L-REF) $\mathsf{Bew}_T(\ulcorner\varphi\urcorner) \to \varphi$.

This is standardly interpreted as a way to codify the *soundness* of T, i.e. the thought that T proves only true sentences. The *global reflection principle*, which we present here in a simplified form, more closely expresses this thought:

(G-REF) $\mathsf{Bew}_T(\ulcorner\varphi\urcorner) \to \mathsf{Tr}(\ulcorner\varphi\urcorner)$.

17. We will return in due course to the question what exactly c' should look like (see p. 9). We should note that requiring that c' strictly extends c is not necessary to block the derivation of the Liar and other paradoxes. In the case of the Liar Paradox, one could simply assume that c' is different from c, at least in that λ expresses a true proposition in c', allowing some other propositions to be in c but not in c'. However, simply assuming that c' is different from c does not sit well with the contextualist approaches of Parsons and Glanzberg, whose theories postulate the existence of an ever expanding hierarchy of contexts. Moreover, the account of context shift to be developed below strongly suggests that contexts expand as a result of uses of necessitation.

18. It might be thought that C-NEC is unnecessarily strong: while it might be plausibly invoked to deal with typically paradoxical sentences, such as λ, it would seem that the truth of the propositions expressed by $0 = 0$ and other 'unproblematic' sentences does not require a context shift. That is, there might be sentences, such as $0 = 0$, that express true propositions in the same context in which they are proved. We are sympathetic to this suggestion, which we aim to explore in future work. In this paper, we stay within the boundaries of the standard contextualist picture, which does not postulate any distinction between 'problematic' and 'unproblematic' sentences. We thank an anonymous referee for useful comments on this point.

19. See e.g. Van Dalen (2003, Chapter 8).

It has been forcefully argued, most notably by Solomon Feferman (1962, 1991) and Georg Kreisel (1970), that the acceptance of a theory T enjoins the implicit acceptance of the soundness of T. On the above understanding of reflection principles, this means that the acceptance of T enjoins the acceptance of (all instances of) a reflection principle for T.

Now for some basic mathematical logic. Even the weakest forms of reflection are known to be unprovable in minimally strong theories, on pain of inconsistency. If T is strong enough to interpret a small amount of arithmetic (and derive L1-L3), then T proves all instances of L-REF only if it is inconsistent. This is, in essence, Löb's Theorem, that if $\vdash_T \mathsf{Bew}_T(\ulcorner\varphi\urcorner) \rightarrow \varphi$, then $\vdash_T \varphi$.[20] On minimal assumptions about the behaviour of the truth predicate, an analogous result holds for G-REF. To be sure, one can consistently add all the instances of L-REF or G-REF for the theory T *to* T. But this yields *a strictly stronger theory*, call it T', that expresses the soundness of T and codifies one's implicit commitment to the soundness of T. The same reasoning that leads to the acceptance of the reflection principles for T can now be used to motivate the acceptance of reflection principles for T'. That is, one is naturally led to accept the theory T'' which results from adding to T' all the instances of L-REF or G-REF for T'. And so on. The process extends in the transfinite (Feferman, 1962).

The idea that expressing the soundness of a theory T requires a strictly stronger theory T' finds a natural articulation in a contextualist framework. As we have seen in §1, contextualist theories come with an explicit indication of the context of reasoning in which a certain derivation takes place.[21] Thus, a 'contextualist theory' is not simply a collection of axioms and rules: it is rather a collection of axioms and rules *in context*. Where Bew^c_S expresses derivability in the theory S_c, this motivates the following contextualist reflection principle:

(C-REF) $\vdash^{c'}_S \mathsf{Bew}^c_S(\ulcorner\varphi\urcorner) \rightarrow \exists_{c'} p(Exp(\ulcorner\varphi\urcorner, p, c') \wedge \mathbf{Tr}(p))$,

for some context c' that strictly extends c.[22] In keeping with the standard treatment of arithmetical reflection rehearsed one paragraph back, we assume that c' is just like c except that its consequence relation $\vdash^{c'}_S$, unlike that of c, contains all instances of C-REF. It is precisely in this sense that c' is stronger than c. More specifically, then, C-REF asserts that provability in the theory S_c is sound in the following sense: if φ is provable in S_c, then φ expresses a true proposition in a stronger context c', i.e. it is proved to express a true proposition in a stronger context c' by a stronger theory $S_{c'}$, whose derivability relation includes all instances of C-REF. That is, just like all the instances of L-REF or G-REF for a non-contextualist theory T are only provable in a strictly stronger theory T', so too all the instances of C-REF that concern Bew^c_S are only provable in a strictly stronger theory $S_{c'}$. Likewise, the arguments for accepting non-contextualist reflection principles such as L-REF or G-REF carry over to C-REF. If one accepts the theory S_c, then one is also committed to its soundness, i.e. to the truth of its sentences.

Now finally to the connection between reflection principles and the Liar reasoning. We first note that C-NEC is an immediate consequence of, and is therefore justified by, C-REF (together with EXT, introduced in §1). To see this, we may reason as follows, where, as above, $\vdash^{c'}_S$ contains all instances of C-REF and c' strictly extends c:

(20) $\vdash^c_S \varphi$ [assumption]

(21) $\vdash^c_S \mathsf{Bew}^c_S(\ulcorner\varphi\urcorner)$ [20, context-relative version of L1]

(22) $\vdash^{c'}_S \mathsf{Bew}^c_S(\ulcorner\varphi\urcorner) \rightarrow \exists_{c'} p(Exp(\ulcorner\varphi\urcorner, p, c') \wedge \mathbf{Tr}(p))$ [C-REF]

(23) $\vdash^{c'}_S \mathsf{Bew}^c_S(\ulcorner\varphi\urcorner)$ [21, EXT]

20. See Boolos (1993, Chapter 3). The standard example to see this in arithmetic is taking φ to be $0 \neq 0$. If we had all instances of L-REF in T, then T would also prove $\neg 0 \neq 0 \rightarrow \neg\mathsf{Bew}_T(\ulcorner 0 \neq 0\urcorner)$; but since T proves $\neg 0 \neq 0$, by *modus ponens* it would also prove $\neg\mathsf{Bew}_T(\ulcorner 0 \neq 0\urcorner)$, which is a formalised statement of the consistency of T, against Gödel's Second Incompleteness Theorem.

21. This is evident from the use of the indexed turnstile $\vdash^c_S$, expressing derivability in the contextualist theory S_c.

22. It follows from our characterisation (§1) that every contextualist theory S_c is recursively axiomatisable. We also assume that a standard provability predicate Bew^c_S can be defined for contextualist theories.

(24) $\vdash_S^{c'} \exists_{c'} p(Exp(\ulcorner\varphi\urcorner, p, c') \land \mathbf{Tr}(p))$ [22, 23, *modus ponens*]

(25) If $\vdash_S^{c} \varphi$, then $\vdash_S^{c'} \exists_{c'} p(Exp(\ulcorner\varphi\urcorner, p, c') \land \mathbf{Tr}(p))$. [20-24, conditional proof]

If C-NEC is derivable from C-REF, and if the latter principle involves a context shift, then it should come as no surprise that C-NEC *also* involves a context shift. More precisely, C-NEC shifts context from c to a context c' whose consequence relation includes all instances of C-REF and that is otherwise exactly like c. We suggest, then, that the Liar in context be understood as involving a context-shifting application of C-NEC, as opposed to NEC, at (15). But then, only (16*) follows from (15), and the Liar Paradox is justifiably blocked.[23]

Three observations are in order. First, the foregoing account of context shift lends support to the standard contextualist treatment of the Liar sentence. One can now prove both that λ does not express a true proposition in k and that it expresses a true proposition in a context k' that strictly extends k. Second, the account provides a *diagnosis* of what goes wrong in the Liar reasoning. On the view on offer, it is essentially a consequence of Löb's Theorem that one can only attribute truth to a proposition expressed by a sentence proved in c in a context that strictly extends c. Since NEC violates this requirement, we are now in a position to *explain why* NEC must be rejected and replaced by C-NEC.[24] Third, we notice that, on the foregoing account, the following weaker rule of necessitation, to the effect that if φ is provable in c, then it expresses a proposition in c' (where c' strictly extends c), is derivable:

(C-NEC$^-$) If $\vdash_S^{c} \varphi$, then $\vdash_S^{c'} \exists_{c'} p(Exp(\ulcorner\varphi\urcorner, p, c'))$.

The rule allows one to justifiably block the derivation of the paradox to be found in Glanzberg (2001, 2004a), where a contradiction is reached by deriving at line 13 that λ doesn't express a proposition and by applying

(NEC$^-$) If $\vdash_S^{c} \varphi$, then $\vdash_S^{c} \exists_{c} p(Exp(\ulcorner\varphi\urcorner, p, c))$

at line 15 to conclude that λ expresses a proposition in c after all. (More precisely, NEC$^-$ yields an inconsistency with line 13 if applied to line 15. See also footnote 13 above.) But of course, NEC$^-$ is unacceptable for the same reasons why NEC cannot be accepted either, and the principle whose use we recommend instead, namely C-NEC$^-$, is justified by C-REF in the same way as C-NEC is.

It might be objected that the foregoing diagnosis is mistaken, on the grounds that the correct contextualist version of G-REF must be

(C-REF*) $\vdash_S^{c'} \mathsf{Bew}_S^{c}(\ulcorner\varphi\urcorner) \to \exists_{c} p(Exp(\ulcorner\varphi\urcorner, p, c) \land \mathbf{Tr}(p))$,

rather than C-REF, for some context c' that strictly extends c. Where C-REF says, in c', that if φ is proved in c then it expresses a true proposition in c', C-REF* says, in c', that if φ is proved in c then it expresses a true proposition *in* c. It might then be maintained that, with C-REF*

23. For a related diagnosis in the context of the Knower Paradox, see Anderson (1983) and Égré (2005, §4).

24. In a recent paper, Glanzberg (2015) distinguishes between truth hierarchies and what he calls *Kreiselian hierarchies*: hierarchies of theories originated by expressing one's acceptance of T via the acceptance of a reflection principle for T. He then suggests that the two may be very different, on the grounds that the addition of a (global or uniform) reflection principle to the Friedman-Sheard theory of truth (Friedman and Sheard (1987), henceforth FS) yields an inconsistent theory, so that one may not expect in general to obtain a truth hierarchy via an iteration of reflection principles for ever stronger truth theories (for global and uniform reflection, and the relative results about FS, see Halbach, 2011, pp. 191-192). Now, it is a consequence of our diagnosis of the Liar that truth hierarchies are very closely tied to Kreiselian hierarchies. However, we don't find the FS example particularly worrying. While we agree that Kreiselian hierarchies don't automatically generate truth hierarchies, we should also note that, as Glanzberg himself observes, the reason why one cannot consistently add a reflection principle to FS is that FS is ω-inconsistent (this is essentially because in FS the truth predicate commutes symmetrically with every logical constant, including negation and the universal quantifier). But since ω-inconsistency is arguably an undesirable feature for a theory of truth, we don't think that the case of FS constitutes a convincing enough reason for thinking that truth and Kreiselian hierarchies in general come apart. For background on FS, see Halbach (2011, Chapter 14).

in place, the Liar Paradox is no longer blocked. The argument would proceed in two steps. One would first show that C-REF* now entails a third version of NEC, viz.

(NEC*) If $\vdash^c_S \varphi$, then $\vdash^{c'}_S \exists_c p(Exp(\ulcorner\varphi\urcorner, p, c) \wedge \mathbf{Tr}(p))$,

for some context c' that extends c. One would then point out that, unlike NEC, NEC* allows for a shift of context in the Liar derivation, but, unlike C-NEC, this does not require that the context parameters occurring in its consequent be indexed to the new context. It might then be argued that NEC* doesn't block the Liar derivation.

Let us examine both steps in turn. The following derivation seemingly establishes the first step:

(20) $\vdash^c_S \varphi$ [assumption]

(21) $\vdash^c_S \mathsf{Bew}^c_S(\ulcorner\varphi\urcorner)$ [20, context-relative version of L1]

(22′) $\vdash^{c'}_S \mathsf{Bew}^c_S(\ulcorner\varphi\urcorner) \rightarrow \exists_c p(Exp(\ulcorner\varphi\urcorner, p, c) \wedge \mathbf{Tr}(p))$ [C-REF*]

(23′) $\vdash^{c'}_S \mathsf{Bew}^c_S(\ulcorner\varphi\urcorner)$ [21, EXT]

(24′) $\vdash^{c'}_S \exists_c p(Exp(\ulcorner\varphi\urcorner, p, c) \wedge \mathbf{Tr}(p))$ [22′, 23′, *modus ponens*]

(25′) If $\vdash^c_S \varphi$, then $\vdash^{c'}_S \exists_c p(Exp(\ulcorner\varphi\urcorner, p, c) \wedge \mathbf{Tr}(p))$. [20-24′, conditional proof]

One can then check that, with NEC* in place, while context is still allowed to shift between lines 15 and 16, the new line 16

(16*) $\vdash^{k'}_S \exists_k p(Exp(\ulcorner\lambda\urcorner, p, k) \wedge \mathbf{Tr}(p))$,

which is derived in k', still yields a contradiction, namely

(17*) $\vdash^{k'}_S \neg\exists_k p(Exp(\ulcorner\lambda\urcorner, p, k) \wedge \mathbf{Tr}(p))$,

which, given EXT, immediately follows from line 13 in the original Liar derivation (the claim, derivable in k, that λ doesn't express a true proposition in k).

The objection is problematic, however. To see this, we appeal to an inverse of the principle of extension introduced in §1, which we call *reducibility*:

(RED) If $\vdash^{c'}_S \varphi$, c extends the context parameters occurring in φ, and c' extends c, then $\vdash^c_S \varphi$.

This principle immediately follows from the rationale behind the use of context parameters in the derivability relation and from the principle C-DER, that if φ is proved in c, then c extends φ's context parameters, put forward in §1. It says that, if c' extends c and c extends the context parameters (if any) occurring in φ, then φ can *also* be proved in c. That is, if φ is derivable in the theory $S_{c'}$ but (i) c already extends all the context parameters in φ and (ii) c is extended by c', then φ is already derivable in the theory S_c. Thus, RED is a kind of parsimony principle: if φ is provable in a theory with context parameter c, the principle guarantees that φ is also provable in the theory with any smaller context parameter that still suffices to extend all the context parameters in φ. Let us now apply RED to C-REF*. We then get the following contextualist reflection principle:

$$\vdash^c_S \mathsf{Bew}^c_S(\ulcorner\varphi\urcorner) \rightarrow \exists_c p(Exp(\ulcorner\varphi\urcorner, p, c) \wedge \mathbf{Tr}(p)).$$

But this is obviously unacceptable, or so Löb's Theorem teaches us: it entails (via a straightforward contextualist variation of Löb's Theorem) that every sentence expresses a true proposition.

We conclude, then, that C-REF* is not a viable interpretation of the contextualist version of the global reflection principle G-REF. In a contextualist framework, G-REF should rather be interpreted as C-REF. But, as we have seen, such a principle entails, and therefore motivates, C-NEC. In turn, C-NEC blocks the Liar derivation and vindicates the basic contextualist claim that context both shifts and expands in the course of the Liar reasoning. Since C-REF is effectively a reflection principle which incorporates, in a contextualist framework, the main lesson of Löb's Theorem, our diagnosis of the Liar Paradox can be summarised in a slogan as the claim that the key for solving the Liar Paradox lies in a proper appreciation of Löb's Theorem in a contextualist framework.

5. Revenge

How robust is the foregoing diagnosis of the Liar? It is often argued that *any* purported solution to the semantic paradoxes faces inevitable *revenge* problems, both in classical and in nonclassical settings.[25] That is, the thought goes, any attempted solution to the semantic paradoxes itself gives rise to Liar-like paradoxes that it cannot solve. We suggest that this is not so in the case of the approach we have just sketched. We consider two possible revenge arguments against contextualist theories, respectively advanced by Andrew Bacon (2015) and (among others) Cory Juhl (1997). We argue that our diagnosis of the Liar Paradox, i.e. the account of context shift we have presented in §4, points to where both arguments go wrong. More precisely, we show that Bacon's and Juhl's paradoxical arguments are effectively versions of the Liar Paradox and that, for this reason, the original contextualist treatment of the original Liar also applies to them.

5.1 A new revenge argument?

According to Bacon, classical approaches that conceive of truth as a property of *sentences* or *utterances* are necessarily subject to a specific kind of revenge paradox—one that, in Bacon's view (p. 306), extends to contextualist theories. Bacon's argument is premised on the assumption that classical approaches must *explain why* certain instances of the T-Schema

(TS) $\varphi \leftrightarrow \mathsf{Tr}(\ulcorner \varphi \urcorner)$,

do not hold. According to Bacon, such an explanation requires distinguishing between *healthy* sentences, to which TS applies, and *unhealthy* ones, to which it doesn't. That is, theorists accepting classical logic are committed to the following restriction of TS:

(SRT) $\mathsf{H}(\ulcorner \varphi \urcorner) \to (\varphi \leftrightarrow \mathsf{Tr}(\ulcorner \varphi \urcorner))$,

25. For an overview on revenge paradoxes, see the essays in Beall (2007). For a generalised form of revenge affecting the main families of non-classical theories of truth, see Murzi and Rossi (2017).

where H is a predicate of (names of) sentences expressing healthiness. Bacon shows, using a sentence ϑ identical to $\mathsf{H}(\ulcorner \vartheta \urcorner) \to \neg\mathsf{Tr}(\ulcorner \vartheta \urcorner)$, that for every classical theory T that interprets Robinson's Arithmetic[26] and contains every instance of SRT, the following holds:

(26) $\vdash_T \vartheta \wedge \neg\mathsf{H}(\ulcorner \vartheta \urcorner)$.

That is, restricting TS to healthy sentences doesn't prevent T from proving of one of its theorems that it is unhealthy. In Bacon's view, this is a paradoxical result, even without assuming that H satisfies a necessitation rule to the effect that if T proves φ, then T also proves $\mathsf{H}(\ulcorner \varphi \urcorner)$. The thought is essentially that no adequate semantic theory should prove theorems that are unhealthy by its own lights.[27]

While Bacon's argument may effectively prove problematic for a range of classical theories, we are not persuaded that it applies equally well to the contextualist approach sketched in §§1-4. The reason is essentially that contextualist approaches are *already committed* to asserting some sentences together with the claim that they do not express propositions in the initial context of reasoning. Since contextualists restrict the T-Schema to sentences that express a proposition (see the principle CTS in §1 above), it follows in Bacon's terminology that contextualists are already committed to asserting sentences that are unhealthy by their own lights. The sentence λ is a case in point. It follows from our proposed treatment of the Liar in context that both λ and the claim that λ doesn't express a proposition in the initial context of utterance k are provable in k. More formally, it can be easily checked that the following sentence is an immediate consequence of (13) and (15) above:

(27) $\vdash^k_S \lambda \wedge \neg\exists_k p(Exp(\ulcorner \lambda \urcorner, p, k))$.

To see that the argument equally applies to (26), we first need to inter-

26. The small amount of arithmetic needed to prove the Diagonal Lemma, that for every formula with one free variable $\varphi(x)$ there is a sentence γ provably equivalent to $\varphi(\ulcorner \gamma \urcorner)$.
27. Similar arguments are also discussed in Reinhardt (1986).

pret $\mathsf{H}(\ulcorner\varphi\urcorner)$ in a contextualist framework. Following Bacon, we identify healthiness with whatever property of sentences is used to restrict the T-Schema. Since contextualists restrict the T-Schema to sentences that express propositions, we say that a sentence is healthy *in c* if it expresses a proposition in *c*. More formally:

(C-H) $\vdash^c_S \mathsf{H}(\ulcorner\varphi\urcorner) \leftrightarrow \exists_c p(Exp(\ulcorner\varphi\urcorner, p, c))$.

On this interpretation, Bacon's paradoxical sentence ϑ is the claim that, if ϑ is healthy, i.e. if it expresses a proposition in k, then it does not express a true proposition in k:

(ϑ) $\exists_k p(Exp(\ulcorner\vartheta\urcorner, p, k)) \rightarrow \neg\exists_k p(Exp(\ulcorner\vartheta\urcorner, p, k) \wedge \mathbf{Tr}(p))$.

A version of Bacon's argument now proves the following contextualist interpretation of (26):

(28) $\vdash^k_S \vartheta \wedge \neg\exists_k p(Exp(\ulcorner\vartheta\urcorner, p, k))$.

But why should (28) be problematic? Line (28) is exactly of the same form as (27): a direct consequence of the Liar in context. Thus, if there is a problem with (28), there is *already* a problem with (27), i.e. with the contextualist approach to the Liar. However, Bacon has not argued for this claim. And it can certainly not be assumed that the contextualist approach to the Liar is defective in an argument aimed at undermining this very approach.

To be sure, it might be argued that (27) and (28) are counterintuitive. After all, they both express the thought that one of S_k's theorems doesn't express a proposition. And how could one trust a theory, if not all of its theorems express propositions? However, contextualists have a standard reply to this objection. In their view, the apparent counterintuitiveness of sentences such as (27) and (28) is made up for at the 'next level', namely in a context that strictly extends k. For instance, as in the Liar in context, classical logic and C-NEC allow one to infer

(29) $\vdash^{k'}_S \exists_{k'} p(Exp(\ulcorner\vartheta\urcorner, p, k'))$

from (28), where k' strictly extends k. That is, contextualist theories can consistently assert that sentences such as λ and ϑ express propositions in k' (in Bacon's terminology: that these sentences are healthy), where k' strictly extends k. As we have seen in §4, this essentially follows from C-NEC, which is in turn motivated by C-REF. We conclude that the contextualist treatment of the Liar described in §4 naturally extends to Bacon's proposed revenge argument and that Bacon has offered no reasons for thinking that either treatment is problematic.

It might be objected that our discussion is not faithful to the letter of Bacon's argument. Bacon suggests that φ is healthy if every utterance of φ is healthy (p. 312), whereas our argument assumes instead that an occurrence of φ in c is healthy if it expresses a proposition in c. However, Bacon's suggested reading still does not give rise to a genuine revenge paradox for the contextualist. Suppose $\mathsf{H}(\ulcorner\varphi\urcorner)$ is interpreted as the claim that every proposition expressed by φ in a context c is healthy. One may then introduce a healthiness operator acting on propositions (which we indicate as **H**, by analogy with truth), and formalise 'φ is healthy' as:

$$\forall_c p(Exp(\ulcorner\varphi\urcorner, p, c) \rightarrow \mathbf{H}(p)).$$

This suggests that Bacon's sentence ϑ be formalised as:

$$[\forall_k p(Exp(\ulcorner\vartheta\urcorner, p, k) \rightarrow \mathbf{H}(p))] \rightarrow [\neg\exists_k p(Exp(\ulcorner\vartheta\urcorner, p, k) \wedge \mathbf{Tr}(p))].$$

Running Bacon's argument in the foregoing modified contextualist framework, one then gets the following conclusion:

(30) $\vdash^k_S \vartheta \wedge \neg\forall_k p(Exp(\ulcorner\vartheta\urcorner, p, k) \rightarrow \mathbf{H}(p)).$

However, it should be clear from the above discussion that (30) is not worrying for the contextualist. The contextualist is already committed to accept sentences that do not express a true proposition in a certain context (this is a direct consequence of (14) and (15) above). Why should the acceptance of sentences that do not always express *healthy* propositions in a given context be any worse?[28]

In his paper, Bacon considers a second possible revenge argument for contextualist approaches. As he puts it, contextualist approaches 'are susceptible to a particularly simple revenge paradox, an instance of our general theorem [(26)]' (p. 313). Such approaches 'are committed to certain instances of the principle that if some utterance of 'φ' is true, then φ' (*Ibid.*). However, a Liar-like argument employing the sentence $\lambda^\star$ identical to 'no utterance of $\lambda^\star$ is true', shows that

> no utterance of $\lambda^\star$ is true, and therefore [the contextualist] should surely be able to assert the fact that no utterance of $\lambda^\star$ is true. Yet when the theorist tries to express this commitment, by uttering the sentence 'no utterance of $\lambda^\star$ is true', he fails. Since according to his own view, *there are no true utterances of the sentence 'no utterance of $\lambda^\star$ is true'—even his very own utterances of this sentence fail to be true, presumably by failing to express a proposition*'. (Bacon, 2015, p. 313; emphasis added; we have adapted Bacon's notation to ours)

In a nutshell, Bacon's point is this. Let us assume the principle that if some utterance of 'φ' is true, then φ.[29] Then, one can establish via Bacon's Liar-like argument both $\lambda^\star$ and the claim that no utterance of $\lambda^\star$ is true, i.e. that $\lambda^\star$ does not express a true proposition. But, Bacon asks, how can the contextualist truly express this thought, if (i) the thought is $\lambda^\star$'s content and (ii) $\lambda^\star$ says that no utterance of $\lambda^\star$ is true?

The contextualist response we sketched two paragraphs back naturally extends to the present case. According to the contextualist, the truth of a sentence is always relative to a context, and the sentence 'no utterance of $\lambda^\star$ is true' really is of the form 'no utterance of $\lambda^\star$ is true *in k*', where k is the utterance context. But then, since $\lambda^\star$ and the claim that no utterance of $\lambda^\star$ is true are proved, and asserted, *in k*, all that follows from the present Liar reasoning is that no utterance of $\lambda^\star$ is true *in k*. That is, in our framework, $\lambda^\star$ does not express a true proposition *in k*. However, as before, an application of C-NEC now yields that $\lambda^\star$ expresses a true proposition in a context k' that strictly extends k. It follows, then, that the conclusion of Bacon's 'particularly simple' revenge paradox (italicised in the quotation in the preceding paragraph) must

28. Bacon may alternatively be interpreted as quantifying over absolutely all contexts, so that a sentence is healthy if every utterance of it, *in any context*, expresses a proposition. On this interpretation, Bacon's argument would in effect be a version of Parsons' 'Superliar', which we discuss in some detail in §5.2 below.

29. This is in effect an utterance-based version of (the right-to-left direction of) CTS, the context-relative version of the T-Schema accepted by contextualists.

be false: it ignores the C-NEC-induced context shift that, according to our version of contextualism, occurs in Liar-like reasonings.

5.2 *Absolute generality and the Superliar*

We now move to a second, relatively standard revenge argument against contextualist approaches. The argument assumes that one can unrestrictedly quantify over absolutely every context, and considers a sentence λ^* which, in a given context k^*, says of itself that it does not express a true proposition *in absolutely any context*. By the now familiar Liar reasoning, λ^* can be shown to be both true and false in absolutely every context. More precisely, having proved λ^* in context c, the application of C-NEC cannot shift context to a *new* context c' that extends c such that λ^* expresses a true proposition in c', since *ex hypothesi* there is absolutely no context in which λ^* expresses a proposition. Parsons (1974, p. 406) calls this version of the Liar the *Superliar*.

Contextualists typically reply to this objection by pointing out that it fails to appreciate that the rejection of absolute generality, understood as the claim that it is possible to quantify over 'absolutely everything', is integral to the contextualist approach to semantic paradox.[30] More specifically, a contextualist would insist that uses C-NEC are context-shifting: if one runs the Liar Paradox in a context c, one ends up in a new context c' whose domain of propositions is strictly greater than that of c. Since this is a perfectly general feature of the view, contextualists reject the possibility of absolutely unrestricted quantification over contexts and propositions. As Glanzberg puts it, the contextualist approach 'require[s] an *open ended* hierarchy of contexts, and [propositions] available in them' (Glanzberg, 2001, p. 240). What is more, the rejection of absolute generality is no *ad hoc* move expressly made in order to ward off revenge. It is rather motivated by the context-shifting properties of C-NEC, which in turn directly follow from the standard lesson of Löb's Theorem applied in a contextualist framework, namely that a theory T cannot prove all instances of its reflection principles.

30. See e.g. Parsons (1974, p. 404) and Burge (1979, p. 192).

Juhl (1997, p. 203) criticises the foregoing contextualist rejoinder, on the grounds that it doesn't apply to what he calls the *Contextual Superliar*.[31] Juhl makes a number of crucial (and controversial) assumptions, in particular that 'whatever contexts are, they collectively form a well-orderable class' (p. 203). For this reason, and in this part of the paper only, we assume that contexts form a well-ordering. Juhl also assumes the existence (and representability in the object-language) of the following function f from sentences to contexts:

> $f(\varphi) = c$ if and only if c is the least context such that φ expresses a true proposition in c.[32]

Juhl then considers the following Liar-like sentence:

(γ) For all contexts c up through $f(\gamma)$, there is no true proposition expressed by γ in c.

In a nutshell, Juhl shows via the usual Liar reasoning that reasoning about γ and $f(\gamma)$ yields a contradiction. We may reconstruct his reasoning as follows. Either γ does not express a true proposition in any context, or it expresses a proposition in some context. Suppose the former. Then, *a fortiori*, γ does not express a true proposition in $f(\gamma)$. But this is what γ says, whence γ must be true in the context immediately succeeding $f(\gamma)$ in the well-ordering of contexts. However, this contradicts our supposition. Now suppose that γ expresses a true proposition in some context. By definition of f, γ expresses a true proposition in $f(\gamma)$. But this contradicts what γ says, namely that there is no context

31. Juhl's criticism is directly targeted against Burge (1979)'s version of the contextualist theory, in which the truth predicate is itself context dependent. In what follows, we adapt Juhl's argument to the contextualist approach sketched in §§1-4. (We also slightly simplify his original argument.)

32. Juhl does not specify the value of $f(\varphi)$ if there is no context in which φ expresses a true proposition. But we can ignore this complication: we might simply stipulate that, in such a case, $f(\varphi)$ returns some conventionally chosen context.

in which γ expresses a true proposition up through $f(\gamma)$. Either way, γ leads to contradiction.

The structure of Juhl's argument is not significantly different from Parsons' Superliar: both presuppose absolutely general quantification over contexts. In particular, the function f is supposed to have the absolute (and well-ordered) totality of contexts as its range: for each given sentence φ, the function f searches amongst 'absolutely all' contexts, and returns the 'least' context in which φ expresses a true proposition. However, contextualists clearly reject the existence of such a function, since it presupposes absolutely unrestricted quantification over contexts. On a contextualist interpretation of f, Juhl's Liar-like reasoning does not yield a contradiction, and γ is essentially treated as the standard Liar sentence λ (or as Bacon's sentence ϑ). In particular, the reasoning yields the following consequences (letting k be the initial context of reasoning, where k extends $f(\gamma)$):

(31) $\vdash^k_S \gamma$ [first half of the Liar reasoning];

(32) $\vdash^k_S$ 'there is no true proposition expressed by γ in any context c in $f(\gamma)$' [definition of γ];

(33) $\vdash^{k^+}_S$ 'there is a true proposition expressed by γ in k^+', where k^+ strictly extends k [C-NEC, second half of the Liar reasoning].

In keeping with their reaction to the standard Liar, contextualists accept both lines 31 and 32, thus maintaining that γ does not express a true proposition in any context in $f(\gamma)$, and line 33, thus maintaining that (given C-NEC) γ expresses a true proposition in a context that strictly extends all the contexts in $f(\gamma)$. As in the original Liar reasoning, (31)-(33) are perfectly consistent and, once again, this is essentially because of the use of C-NEC in the course of the paradoxical derivation.

Notice that we *must* take as the initial context of reasoning a context k that extends $f(\gamma)$. This is a direct consequence of C-DER, the principle that if φ is proved in c, then c extends φ's context parameters. That is, C-DER forces line 31 to be proved either in $f(\gamma)$ or any context that extends $f(\gamma)$. If we took a context that did not extend the contexts mentioned in γ (and $f(\gamma)$ is one such context), we would end up with a proof of γ in a context that is strictly extended by $f(\gamma)$, thus violating C-DER and the rationale behind the use of context-relative derivability relations. We conclude, then, that once we interpret f in a way that is acceptable for the contextualist, the machinery put in place to address the standard Liar Paradox applies equally well to Juhl's version of the Superliar.[33]

6. Concluding remarks

Gauker writes that

> We cannot maintain that if a sentence φ is provable from sentences that are true in some context, then φ must express a proposition in some *other* context. The most we can maintain is that if a sentence φ is provable from premises that are true in a given context, then φ must express a proposition in that same context. (Gauker, 2006, p. 403)

We hope to have shown that Gauker's pessimism is misplaced. The Liar reasoning involves the application of a principle, C-NEC, which is derived from, and hence justified by, a reflection principle C-REF to the effect that sentences that are proved in a given context c express true propositions in a context c' that strictly extends c. The context-shifting properties of C-REF are motivated by well-known facts about reflection principles for arithmetic, and properties of contexts. In turn, the context-shifting properties of C-NEC are inherited by those of C-REF, from which C-NEC follows. And since the context-shifting properties of C-REF simply are a consequence of Löb's Theorem in a contextualist framework, our basic diagnosis of the Liar is that the invalidity of the Liar reasoning is ultimately a consequence of Löb's Theorem. We have

33. Elia Zardini (2008, §4) discusses yet more Superliar-like puzzles. However, all of them can be interpreted as variants of the ordinary Liar Paradox and hence blocked along the lines we have just sketched in our discussion of Parsons' and Juhl's paradoxes.

also suggested that our diagnosis is robust. Once the Liar in context is construed as involving uses of C-NEC, certain revenge arguments directed at contextualist approaches are treated exactly as the original Liar. Contextualist approaches may face a number of difficulties.[34] But explaining why context shifts in the course of the Liar reasoning need not be one of them.[35]

34. See e.g. Priest (2006, §§1.5-6) and Field (2008, Chapter 14).

35. We wish to thank the participants of the philosophy of logic reading group in Sheffield (especially Dominic Gregory and Jonathan Payne), Christopher Gauker and Michael Glanzberg for helpful exchanges about the Liar over the years, and Paul Egré, Christopher Gauker, Dave Ripley, Gil Sagi, Lionel Shapiro, two anonymous referees, and the editors of *Philosophers' Imprint* for comments on earlier drafts that have led to considerable improvements. We are grateful to the ÖAD and the FWF (grant number P2971-G24) for generous financial support during the time this paper was written.

References

Anderson, C. A.: 1983, The Paradox of the Knower, *Journal of Philosophy* **80**(6), 338–355.

Bacon, A.: 2015, Can the classical logician avoid the revenge paradoxes?, *Philosophical Review* **124**(3), 299–352.

Beall, J. (ed.): 2007, *Revenge of the Liar*, Oxford University Press, Oxford.

Beklemishev, L. D.: 2005, Reflection principles and provability algebras in formal arithmetic, *Russian Mathematical Surveys* **60**, 197–268.

Boolos, G.: 1993, *The Logic of Provability*, Cambridge University Press, New York and Cambridge.

Burge, T.: 1979, Semantical paradox, *Journal of Philosophy* **76**(4), 169–98.

Chihara, C.: 1979, The semantic paradoxes: A diagnostic investigation, *The Philosophical Review* **88**(4), 590–618.

Égré, P.: 2005, The Kower Paradox in the light of provability interpretation of modal logic, *Journal of Logic, Language, and Information* **14**(1), 13–48.

Feferman, S.: 1962, Transfinite recursive progressions of axiomatic theories, *Journal of Symbolic Logic* **27**, 259–316.

Feferman, S.: 1991, Reflecting on Incompleteness, *Journal of Symbolic Logic* **56**(1), 1–49.

Field, H.: 2008, *Saving Truth from Paradox*, Oxford University Press, Oxford.

Friedman, H. and Sheard, M.: 1987, An axiomatic approach to self-referential truth, *Annals of Pure and Applied Logic* **33**, 1–21.

Gauker, C.: 2006, Against stepping back: A critique of contextualist approaches to the semantic paradoxes, *Journal of Philosophical Logic* **35**(4), 393–422.

Gauker, C.: 2010, Global domains versus hidden indexicals, *Journal of Semantics* **26**, 243–270.

Glanzberg, M.: 2001, The liar in context, *Philosophical Studies* **103**(3), 217–51.

Glanzberg, M.: 2004a, A contextual-hierarchical approach to truth and the liar paradox, *Journal of Philosophical Logic* **33**, 27–88.

Glanzberg, M.: 2004b, Truth, reflection, and hierarchies, *Synthese* **142**(3), 289–315.

Glanzberg, M.: 2015, Complexity and Hierarchy in Truth Predicates, *in* T. Achourioti, H. Galinon, J. Martínez Fernández and K. Fujimoto (eds), *Unifying the Philosophy of Truth*, Vol. 36 of *Logic, Epistemology, and the Unity of Science*, Springer.

Halbach, V.: 2011, *Axiomatic Theories of Truth*, Cambrdige University Press, Cambridge.

Heck, Jr., R. G.: 2007, Self-reference and the languages of arithmetic, *Philosophia Mathematica* **15**, 1–29.

Heck, Jr., R. G.: 2012, A liar paradox, *Thought* **1**, 36–40.

Heck, Jr., R. G.: 2013, More on 'a liar paradox', *Thought* **1**, 270–280.

Juhl, C. F.: 1997, A context-sensitive liar, *Analysis* **57**(3), 202–204.

Kaplan, D.: 1989, Demonstratives, *in* J. Almog, J. Perry and H. Wettstein (eds), *Themes from Kaplan*, Oxford: Oxford University Press.

Kreisel, G.: 1970, Principles of proof and ordinals implicit in given concepts, *in* A. Kino, J. Myhill and R. E. Vesley (eds), *Intuitionism and Proof Theory*, North-Holland, Amsterdam, pp. 489–516.

Murzi, J. and Rossi, L.: 2017, Generalised revenge. Unpublished manuscript.

Parsons, C.: 1974, The Liar Paradox, *Journal of Philosophical Logic* **3**(4), 381–412.

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

Reinhardt, W. N.: 1986, Some remarks on extending and interpreting theories with a partial predicate for truth, *Journal of Philosophical Logic* **15**, 219–51.

Rumfitt, I.: 2014, Truth and meaning, *Proceedings of the Aristotelian Society Supplementary Volume* **LXXXVIII**, 21–55.

Scharp, K.: 2013, *Replacing Truth*, Oxford University Press, Oxford.

Stanley, J. and Szabo, Z.: 2000, On quantifier domain restriction, *Mind and Language* **15**, 219–61.

Van Dalen, D.: 2003, *Logic and Structure*, Springer-Verlag, Berlin.

Zardini, E.: 2008, Truth and what is said, *Philosophical Perspectives* **22**, 545–574.