DIALETHEISTS' LIES ABOUT THE LIAR

JONAS R. BECKER ARENHART
Departamento de Filosofia, Universidade Federal de Santa Catarina, BRAZIL
jonas.becker2@gmail.com

EDERSON SAFRA MELO
Departamento de Filosofia, Universidade Federal do Maranhão, BRAZIL
edersonsafra@yahoo.com.br

Abstract.   Liar-like paradoxes are typically arguments that, by using very intuitive resources of natural language, end up in contradiction. Consistent solutions to those paradoxes usually have difficulties either because they restrict the expressive power of the language, or else because they fall prey to extended versions of the paradox. Dialetheists, like Graham Priest, propose that we should take the Liar at face value and accept the contradictory conclusion as true. A logical treatment of such contradictions is also put forward, with the Logic of Paradox (LP), which should account for the manifestations of the Liar. In this paper we shall argue that such a formal approach, as advanced by Priest, is unsatisfactory. In order to make contradictions acceptable, Priest has to distinguish between two kinds of contradictions, internal and external, corresponding, respectively, to the conclusions of the simple and of the extended Liar. Given that, we argue that while the natural interpretation of LP was intended to account for true and false sentences, dealing with internal contradictions, it lacks the resources to tame external contradictions. Also, the negation sign of LP is unable to represent internal contradictions adequately, precisely because of its allowance of sentences that may be true and false. As a result, the formal account suffers from severe limitations, which make it unable to represent the contradiction obtained in the conclusion of each of the paradoxes.

Keywords: Contradiction • Liar paradox • dialetheism • logic of paradox

> Overall, such paradoxes as the Liar provide some evidence for the dialetheist's claim that some contradictions are *provably* true, in the sense that they are entailed by plain facts concerning natural language and our thought processes.
>
> Priest and Berto 2013, sect.3.2

1. Introduction

Semantic paradoxes feature as one of the main evidences to the view that some contradictions are true, *i.e.*, that there are some sentences α, such that both α and $\neg\alpha$

are true. True contradictions, in the sense just defined, are called *dialetheias*, and the defenders of the view are called *dialetheists*. One of the consequences of dialetheism is that some sentences are both true and false, that is, they bear what are now called *truth-value gluts*. These definitions are standard in the literature, and so we shall adopt them throughout the paper (see Priest and Berto (2013) for standard definitions).

Among the semantic paradoxes, the Liar has clear prominence. Dialetheists view the Liar as a valid and sound argument; its premises are true and its contradictory conclusion is true. Since dialetheism is the view that *some* contradictions are true, but *not all* of them, a paraconsistent logic is required in order to grant that contradictions do not imply everything; that is, dialetheists reject the so-called *explosion rule*: $\alpha, \neg\alpha \vdash \beta$. Graham Priest advances his famous Logic of Paradox (LP) as the most appropriate formalism to represent true contradictions without triviality (Priest 1979; 2006a, chap.1). In LP true contradictions bear gluts as their truth-value.

The version of dialetheism we shall consider in this paper, — the one advocated by Priest[1] —, countenances that there are two basic ingredients for an adequate treatment of paradoxes such as the Liar. *The first one* concerns the intuitive resources involved in deriving the paradoxes: the inferences and natural steps taken from simple assumptions that, according to dialetheists, must be preserved in any account of the paradoxes. In particular, the idea that contradictions should be tolerated in the cases of such paradoxes requires that it is precisely *contradictions in the intuitive (or pre-theoretic) sense of the Liar*, as it is framed in natural languages, that must be accounted for. *The second condition* concerns the adequacy of the formal machinery employed to grant that contradictions are faithfully represented and that explosion will not be a threat. What is required is that the negation sign of such a logic faithfully represents contradictions in the informal sense displayed by the Liar and related paradoxes (Priest 2006b, chap.4).

The *first requirement* codifies a constant claim in the works of dialetheists. An advantage of dialetheism, it is said, is that it is the only treatment of paradoxes preserving the very intuitive line of reasoning that gives rise to the Liar (and here we shall restrict our attention to the Liar). According to dialetheists, the Liar is the natural outcome of simple facts concerning natural language and the workings of some of our natural reasoning principles (see Priest and Berto 2013, sect.3.2). It is our informal notion of truth that is clearly inconsistent, and dialetheism seeks to preserve such inconsistency without having to give up any feature of truth as it is pre-theoretically understood. Dialetheists frequently claim that every *consistent* solution advanced so far requires giving up some intuitive principle of natural language or else some natural feature of truth. According to Priest, we have to keep those natural principles untouched and allow contradictions to arise; contradictions are not the problem, it is rather triviality that must be avoided (Priest 2002, p.350).

What the *second requirement* adds to this one is that we must have an adequate paraconsistent logic able to faithfully represent those intuitive (inconsistent) features. Intuitively, the Liar displays a *switch of truth values*: when we assume that the Liar sentence has one of the two possible truth values, then we are led to accept that it really has the other truth value too. So, from an intuitive point of view, it seems that the contradiction is reached precisely when such a switch of truth values takes place. Consistency demands that a single truth value be attributed to a sentence, it seems, or at least that it be not both true and not-true. In this sense, we have a contradiction in the end of the Liar. Due to the assumed soundness of the derivations, dialetheists argue that the Liar argument shows the existence of true contradictions (Priest 2006b, p.83). A paraconsistent logic coping with dialetheists' purposes must represent precisely those features.

In this paper, we shall take issues with the alleged adequacy of the formal counterpart of contradictions in Priest's LP. We shall consider very closely how formulas representing contradictions in this system are related with the pre-theoretic notion of contradiction, and, most important, how they are related to two versions of the Liar: the simple Liar ("this sentence is false") and the extended Liar ("this sentence is not true"). Priest's approach requires a split of contradictions into two categories: "internal contradiction" and "external contradiction". In his account, the simple Liar turns out to be a case of *internal contradiction* and the extended Liar turns out to be a case of *external contradiction*. As we shall see, the two kinds of contradictions are not equivalent, even by dialetheists' standards (see Priest 2006a, sect.4.9); as a result, we shall argue that this version of dialetheism faces problems. The distinction between two kinds of contradictions is required if Priest's dialetheism is to be true. As a result of such a split, however, we can hope that at most one kind of contradiction will be adequately represented formally, which is bad news for a complete treatment of paradoxes. In order to have both kinds of contradictions equivalent, their distinction must collapse, which in its turn is bad news for the truth of dialetheism. Both options are hard to swallow for a dialetheist.

As a further consequence of the preceding difficulties, we argue that dialetheists cannot satisfy both desiderata posed before. We shall see that if we accept that we have faithfully modeled the intuitive reasoning leading to the Liar, delivering thus a contradiction, then it will be impossible to avoid explosion. On the other hand, to avoid explosion and have a legitimate paraconsistent logic, one must allow that the negation sign does not represent contradictions in an intuitively robust sense, precisely the sense that appears in informal formulations of the Liar. In particular, we shall provide new arguments to reinforce the charge that paraconsistent negation is not a contradictory-forming operator. Taking this into account, the dialetheist ends up in a crossroad: to keep the paraconsistent logic and not really have contradictions or else to allow contradictions and lose the non-explosive character of her negation.

Obviously, any path will be harmful to dialetheism. Ultimately, this dilemma imposes restrictions on expressive resources of the dialetheist's approach, an issue that is dear to dialetheists. Thus, we shall argue that a paraconsistent logic like LP cannot accommodate the Liar and that dialetheism cannot offer a genuine answer to the problem of the Liar.

The structure of this paper is as follows. In section 2 we briefly present what some dialetheists refer to as the "essence of the Liar", in order to analyze the intuitive (or pre-theoretic) notion of contradiction that is encapsulated in its conclusion. Then, section 3 has three subsections. In the first we briefly present the traditional formal approach to capture the notion of contradiction, in which the two versions of the paradox are equivalent. In the second one, we present Priest's formal approach to the paradoxes, separating two kinds of contradictions. The third subsection challenges the plausibility of the dialetheist's approach taking into account mainly the extended Liar. In the section 4, we focus on the dialetheist account of negation as a contradictory-forming operator, mainly in the context of the simple Liar. We argue that the paraconsistent logic adopted by dialetheism does not capture, or model, the notion of contradiction and, therefore, there is a mismatch between dialetheism and the underlying logic that presumably would be able to capture the notion of contradiction. Section 5 contains some conclusive remarks.

2. Contradiction as the "essence of the Liar"

In this section, we consider the Liar argument from an informal point of view in order to shed light on the concept of contradiction that is in focus. The Liar derivation delivers a contradiction (in a pre-theoretic sense) in natural language. We have resources to define the Liar sentences and the required reasoning principles for developing such arguments. In the center of the stage is the behavior of the truth predicate. As we know, English and other natural languages can express their own semantic concepts — natural languages are *semantically closed* in the well known Tarskian sense: they have their expressions, names for their expressions and semantic terms to refer to their own expressions.[2]

The T-schema, governing the behavior of the truth predicate, plays an important role in the reasoning leading from the Liar sentence to a contradiction. For every sentence α in the language, the T-schema is as follows:

$$T(\ulcorner\alpha\urcorner) \leftrightarrow \alpha$$

Here T is the truth predicate and the corners are a name-forming device.

Now, employing self-reference, another feature of natural language, we can easily construct the *Liar sentence*. We begin with the *simple Liar sentence*, which asserts of itself that it is *false*.

(SLS) (SLS) is *false*.

Let us reason by cases, assuming the principle of bivalence, according to which all sentences are either true or false. If SLS is true, then it is false; if SLS is false, then it is true (following from the T-schema). Therefore, the SLS is true if and only if it is false. According to dialetheists, the argument shows that SLS is both true and false, *i.e.* it bears a truth-value *glut*, in the dialetheists' terminology (which we presented in the previous section).

Intuitively, it appears that it is difficult to accept that a sentence is both true and false. Thinking of a typical competent English speaker, not specialized in this discussion, it is likely that she would consider that it is more reasonable to give up bivalence, and to assume that the SLS is neither true, nor false (*truth-value gaps*), than to assume that SLS is true and false. However, as is well known, dialetheists have an argument against this move. Using the resources of natural language, such a strategy is bound to fail because we can easily arrive at paradox again, which, in turn, is achieved through the *extended Liar sentence*, which asserts of itself that it is *not true*.

(ELS) (ELS) is *not true*.

Taking excluded middle and T-schema, an argument similar to the simple Liar leads us to conclude that ELS is true if and only if it is not true, which leads to the claim that it is true and not true. This is a contradiction! According to dialetheists, the argument is valid, although it ends with a contradiction.

The extended Liar is a particular case of a general dialetheist revenge argument against consistent solutions to the Liar. It doesn't matter which concept (neither true nor false, nonsense, paradoxical, and so on) is adopted to avoid the contradiction, because in the new context we can use the very same concept of the "solution" to arrive at contradictions again.[3] In order to present the argument in full generality, Priest distinguishes between the *Bona fide* truths and its complement, called *the Rest*. On that basis, Priest lays out what that he calls the "essence of the Liar":

> "the essence of the liar paradox is a particular twisted construction which forces a sentence, if it is in the Bona fide truths, to be in the Rest (too); conversely, if it is in the Rest, it is in the Bona fide truths" (Priest 2006a, p.23).

Berto (2007, sect.2.5) supports Priest's considerations about the twisted construction that keeps the Liar switching from truth to its complement and vice versa, and also goes as far as dubbing this movent "the essence of the Liar". Indeed, it seems that these dialetheists' remarks capture the core of the Liar (or its "essence", to use Priest's and Berto's term).[4] The bite of the argument lies in the fact that if this twisted

movement could be stopped, so that the sentence happens to fall exclusively under the truth or under its complement, there is no longer a Liar (since it would have lost its "essence"). We would no longer have a Liar sentence in this case. Following Priest, in the case of the simple Liar, the complement coincides with the false. If we try to solve the problem admitting sentences that are neither true nor false (as stated above), "being false" then becomes just a proper part of the complement of truth (the other part being sentences neither true nor false, the gaps). But this strategy, according to Priest, does not pose the problem appropriately, for if the false is only a proper part of the complement, then the simple Liar is not the correct formulation of the problem, as Priest advocates (see Priest 2006a, p.23).

However, when the essence of the Liar is lost, as in such gap approaches, the movement of revenge appears to remind us that this supposed stability (taking a position exclusively in the *Bona fide* truth or in its complement) is an illusion; the contradiction always returns, evidenced precisely when the legitimate switch of truth values is re-established. Thus, it is clear that, regardless of the Liar's version being considered (the simple one or some extended version), ultimately the contradiction will always be present, precisely manifested in this switch of truth values — which arguably characterizes the paradox. As Priest says (2006a, p.23), the extended Liar shows us that the terms of a category do not coincide with the complement (*e.g.* falsity and gap); hence, we can return to the original problem by describing the complement in some alternative way (*e.g.* not true).[5] Thus, Priest claims that "the extended paradoxes are not really novel paradoxes, but merely manifestations of *one and the same problem*, suitable to different contexts" (Priest 2006a, p.24, emphasis added). "These paradoxes [extended paradoxes] are just *the same old paradox* in a new theoretical context" (Priest 2007, p.173, emphasis added).[6]

Both SLS and ELS have true contradictions as conclusions, given Priest's remarks on the paradoxes. Therefore, following the dialetheists' approach, ultimately gluts are equivalent to dialetheias. As Priest puts it "(. . .) paradoxical sentences are both true and false — as it is sometimes put, they are semantic gluts (dialetheias)" (Priest 1995, p.57). Bearing in mind these considerations about Liar sentences, gluts, and contradictions, in the next section we discuss their formalization and how they may be seen to present troubles even for a dialetheist.

3. Troubles modeling "the" contradiction

3.1. The traditional formal version

Given that the Liar is primarily an argument couched in natural language, the core of the question for dialetheists is precisely the consistency of the concepts of natural language involved in the derivation.[7] Thus, the dialetheists' aim is not to develop a

formal system restricting some of the informal principles leading to contradiction; the dialetheists' attitude, facing the paradox, is rather to accept the contradiction and provide a suitable formalism able to faithfully model the intuitive reasoning that naturally ends up in contradiction. As Priest (2006a, p.9) puts it:

> What is at issue is the consistency of the familiar concepts which give rise to the paradoxes, or, what comes to the same thing, the consistency of the semantics of fragments of natural language. For example, we may set up a theory in a formal language containing the words 'is true', and this may be consistent. However, the crucial question remains: how adequate a formalisation is this of the phenomenon we are trying to model: natural reasoning?

The emphasis should fall on the adequacy of the formalization, according to Priest. Given that we cannot demand consistency from dialetheists (due to the fact that they accept some contradictions), we must demand that they should at least be coherent, since Priest makes it clear that inconsistency does not entail incoherence (Priest 2006a, p.6). We can tie this coherence, as Priest does in the above quote, to the adequacy between the dialetheists' intuitive considerations and their formal account of it. We consider adequacy to be a mark of coherency, so that in the case we are considering, coherence demands that Priest's formalism must be able to characterize the "essence" of the "same old paradox". Thus, it seems that it doesn't matter which is the complement adopted ("false", "not true", or anything else), since switching from truth to its complement and vice-versa is to be retained (*i.e.*, the contradiction remains).

In the next subsections we shall put some pressure on the coherency of Priest's account. In order to do so, we now present a formalization of the paradoxes, in order to compare how the dialetheist's approach departs from the traditional one and how it fares according to the demands of dialetheists' themselves, as stated in section 1. Suppose that we have a formal theory containing all the machinery of arithmetic, whose language is able to express its own syntax. Let us suppose also that it has a predicate $F(x)$ whose extension is the set of all the false sentences of this theory. The diagonal lemma brings in self-reference. In particular, having $F(x)$, we have a sentence λ such that:

$$\lambda \leftrightarrow F(\ulcorner \lambda \urcorner)$$

Intuitively, this sentence says of itself that it is false. We can say that in this case we have a formal counterpart of SLS.

Considering that falsity is the truth of negation — $F(\ulcorner \alpha \urcorner) \leftrightarrow T(\ulcorner \neg \alpha \urcorner)$ — we have that:

$$\lambda \leftrightarrow T(\ulcorner \neg \lambda \urcorner)$$

Through the T-schema, we have

$$T(\ulcorner\lambda\urcorner) \leftrightarrow T(\ulcorner\neg\lambda\urcorner)$$

Assuming the principle of bivalence — BIV: $T(\ulcorner\alpha\urcorner) \vee T(\ulcorner\neg\alpha\urcorner)$ — we have that $T(\ulcorner\lambda\urcorner) \wedge T(\ulcorner\neg\lambda\urcorner)$; *i.e.*, λ is *truth-value glut*.

The argument using a sentence λ which is equivalent to $\neg T(\ulcorner\lambda\urcorner)$ (a formal counterpart of ELS) is similar. The difference is that the argument uses *the law of excluded middle* (LEM) and ends up in $T(\ulcorner\lambda\urcorner) \wedge \neg T(\ulcorner\lambda\urcorner)$; that is, we end up with an explicit contradiction:[8] true and not true.

As we know, in classical logic LEM is equivalent to BIV. Taking the equivalences $T(\ulcorner\alpha\urcorner) \leftrightarrow \alpha$ and $\neg T(\ulcorner\alpha\urcorner) \leftrightarrow T(\ulcorner\neg\alpha\urcorner)$, the equivalence between LEM and BIV holds. Thus, in classical logic, the equivalence between falsity and untruth holds and, therefore, the equivalence between SLS and ELS holds as well. In the end, due to this equivalence, it doesn't matter whether we use SLS or ELS, the argument of the Liar drives us to a contradiction as expected.[9] Hence, from the formal point of view, using both SLS (*i.e.* a sentence λ, of the form $T(\ulcorner\neg\lambda\urcorner)$) and ELS (*i.e.* a sentence λ, of the form $\neg T(\ulcorner L\urcorner)$), we have a switch between truth values just as we have in the pre-theoretic case.

3.2. The dialetheist's formal version

As we have briefly seen, the simple Liar argument (using SLS) and the extended Liar argument (using ELS) are equivalent in classical logic. This happens mainly due to the fact that $T(\ulcorner\neg\alpha\urcorner)$ is logically equivalent to $\neg T(\ulcorner\alpha\urcorner)$; that is, in classical logic the negation sign may equivalently be put inside the T predicate or else outside it. As a result, the two versions of the paradox are both equally *disastrous* in this case. On the other hand, these arguments should be equally *good* for dialetheists, who wish to accept the resulting contradiction, and through that acceptance "solve" the problem. However, despite its advertisement, Priest's formal approach does not naturally match its own intuitive demands. To advance this claim is the main concern of this and of the next subsection. As we shall argue, there is a mismatch between the intuitive point of view and the formal one in Priest's dialetheism. According to Priest's account, *from an intuitive point of view*, SLS and ELS lead to the "the same old paradox"; however, *from a formal point of view*, when an appropriate paraconsistent logic substitutes classical logic in the formalization of the paradoxical arguments, SLS and ELS are different, since the equivalence between $T(\ulcorner\neg\alpha\urcorner)$ and $\neg T(\ulcorner\alpha\urcorner)$ does not hold according to Priest.

The main ingredient in separating the two versions of the paradox comes from Priest's (2006a, p.70) separation of that equivalence into two implications:

Exhaustion: $\neg T(\ulcorner \alpha \urcorner) \rightarrow T(\ulcorner \neg \alpha \urcorner)$

Exclusion: $T(\ulcorner \neg \alpha \urcorner) \rightarrow \neg T(\ulcorner \alpha \urcorner)$

Exhaustion claims that if a sentence is not true, then it must be false. So, truth and falsity exhaust the possibilities and, as a consequence, it prohibits gaps.[10] *Exclusion*, on the other hand, claims that if a sentence is false, then it is not true. Truth and falsity are mutually exclusive and, as a consequence, gluts are prohibited by exclusion.

In the gap approach exhaustion doesn't hold precisely to grant the existence of gaps (sentences neither true, nor false). So, as we have seen above, SLS and ELS are also not equivalent in a gap approach. According to Priest, a gap approach does not present a correct formulation of the Liar, because as we have seen, when we allow the possibility of gaps, false turns out to be just a proper part of the complement of truth. Consequently, we do not have a contradiction (see section 2 above), and not having a contradiction (gap avoids this, at least in the first analysis), we do not have the Liar. According to Priest, truth and falsity are mutually exhaustive; the mere failure of truth of a sentence is enough for the truth of its negation (see Priest 2006a, sect.4.7).

Now, focusing on exclusion, Priest claims that the question of whether or not truth and falsity are mutually exclusive *is the question of whether dialetheism is true* (Priest 2006a, p.67). According to Priest, semantic paradoxes lead us to dialetheism showing that some false sentences are true and, therefore, that exclusion fails. As was to be expected, Priest's conclusion is that truth and falsity are not mutually exclusive; to dialetheists like him, the Liar is displaying that the intersection of truth and falsity is nonempty (*i.e.*, there are truth-value gluts).

However, the failure of exclusion provides not only for the possibility of truth and falsity to overlap, but also provides for a difference between falsity and untruth. A false sentence may still be true (in case it is a glut), but nothing so far is said about whether a sentence that is untrue can also be true. As we shall see, here lies the crux of the matter when it comes to deal with the conclusion of the extended Liar. But how relevant is that difference for a dialetheistic account of the paradoxes? According to Priest, there is no special importance to be attached to that difference:

> Falsity and untruth are therefore distinct, and if this is so the next question is what significant differences there are between them. The answer is 'surprisingly little'. In particular, truth and untruth are exhaustive and nonexclusive, just as truth and falsity are (Priest 2006a, p.71).

Recall that falsity and untruth are equivalent for the classical logician. The simple and extended versions of the Liar are equivalent in this case. The same does not hold for Priest's dialetheism, though, given the failure of exclusion. The consequences of this will be a strong evidence against this version of dialetheism and the adequacy of

its treatment of the paradoxes. We begin by presenting how to distinguish between the two versions of the Liar paradox.

Starting from the simple Liar, which has as conclusion a sentence true and false, $T(\ulcorner\lambda\urcorner) \wedge T(\ulcorner\neg\lambda\urcorner)$, the exclusion principle gives us a sentence which is true and not true, $T(\ulcorner\lambda\urcorner) \wedge \neg T(\ulcorner\lambda\urcorner)$ (that is exactly the conclusion of the extended Liar). Having the exclusion principle, negation goes from inside to outside of the truth predicate. According to Priest, with exclusion an "internal contradiction" generates an "external contradiction".[11] Without exclusion, the possibility of going back and forth between falsity and untruth is lost, a difference between falsity, $T(\ulcorner\neg\lambda\urcorner)$, and untruth, $\neg T(\ulcorner\lambda\urcorner)$, emerges, and we must acknowledge two kinds of contradictions. While the conclusion of the simple Liar, $T(\ulcorner\lambda\urcorner) \wedge T(\ulcorner\neg\lambda\urcorner)$, is a case of *internal contradiction*, the conclusion of the extended Liar, $T(\ulcorner\lambda\urcorner) \wedge \neg T(\ulcorner\lambda\urcorner)$, is a case of *external contradiction*. Internal contradictions, the sentences that are true and false, are cases of gluts. For external contradictions, as we shall see, this is not so clearly the case.

This difference in two kinds of contradictions infects the two versions of the Liar directly. Let us check how they impact on differences between simple and extended Liar, according to Priest himself. According to Priest's argument, from the conclusion of the extended Liar, $T(\ulcorner\lambda\urcorner) \wedge \neg T(\ulcorner\lambda\urcorner)$, we can derive $\exists x(Tx \wedge \neg Tx)$. Furthermore, from the law of excluded middle and universal generalization, $\forall x(Tx \vee \neg Tx)$ follows, and by de Morgan and quantifier principles, we derive $\neg\exists x(Tx \wedge \neg Tx)$. All of that notwithstanding, from the conclusion of the simple Liar, $T(\ulcorner\lambda\urcorner) \wedge T(\ulcorner\neg\lambda\urcorner)$, we can derive $\exists x(Tx \wedge Fx)$, but its corresponding negation, $\neg\exists x(Tx \wedge Fx)$, cannot be derived without the exclusion principle. Taking this into account, Priest claims that "truth and untruth are, therefore, "more inconsistent" than truth and falsity" (Priest (2006a), p.72). Following Priest's remarks, the extended Liar is "more inconsistent" than the simple Liar. Affirming $\exists x(Tx \wedge \neg Tx)$ and $\neg\exists x(Tx \wedge \neg Tx)$, Priest is admittedly asserting that a contradiction (specifically, an external contradiction, in his classification) does no harm, since dialetheists intend to accommodate contradictions without explosion (Priest 2006a, p.72). As we shall see, there are problems with this claim.

The first, immediate worry, is that by providing for two versions of the Liar there is strong evidence that there will be a problem for representing a contradiction in at least one case of the Liar; we miss the natural flow from the intuitive Liar story to its formal representation. As we have seen, usually the dialetheist's intuitive story begins with the simple Liar leading us to a sentence that is true and false (internal contradiction), a *glut* by definition. Revenge, on the other hand, drives us to a sentence that is true and non-true (external contradiction). According to Priest, in an intuitive context, the simple Liar and the extended ones are "merely the manifestation of one and the *same* problem" (both arguments are sound in natural language). However, in the context of Priest's formal account, this seems not to be the case. The conclu-

sion of the extended Liar, $T(\ulcorner\lambda\urcorner) \wedge \neg T(\ulcorner\lambda\urcorner)$, does not by itself grant that λ is a glut. Someone could urge that the conclusion of the extended Liar *implies* that $T(\ulcorner\lambda\urcorner) \wedge T(\ulcorner\neg\lambda\urcorner)$, showing that λ is, after all, a glut. However, even if this move could be made (we will argue shortly that this interpretation of what is going on can be resisted), we cannot go from $T(\ulcorner\lambda\urcorner) \wedge T(\ulcorner\neg\lambda\urcorner)$ to $T(\ulcorner\lambda\urcorner) \wedge \neg T(\ulcorner\lambda\urcorner)$, due to the banning of exclusion. So, failing the equivalence between $T(\ulcorner\neg\lambda\urcorner)$ and $\neg T(\ulcorner\lambda\urcorner)$, we do not have the equivalence between the conclusions of the simple Liar and the extended Liar; they are not the same paradox. Since the suppression of the exclusion principle splits the contradiction in two different categories, it breaks down the desirable intuitive equivalence between gluts and contradictions.[12]

Under these considerations, unlike the case in the intuitive context, in Priest's formal story simple Liar and extended Liar have distinct properties; they are *not* manifestations of one and the same old problem.[13] In other words: intuitively, the conclusions of both versions of the Liar are supposed to be a contradiction. They are the conjunction of a sentence that is true with its alleged *Rest*. However, the Rest is differently characterized in the different versions of the paradox in the dialetheist's story. In the simple one, it is a false sentence; in the extended one, it is a untrue sentence. Which one is the legitimate Rest? Given that the Rest is agreed to be unique, the trouble for the dialetheist representation is that now there are two incompatible ways to characterize it. On the one hand, the revenge should allow one to go from truth to untruth; but then we seem unable to claim that the conclusion of the extended Liar is a glut (*viz.* true and false). On the other hand, on the case of the simple Liar, the conclusion is a glut, but if we identify gluts with contradictions, there is no clear story to be told of the opposition between truth and untruth. Is that also a contradiction? There is something missing in this picture. As we shall now argue, the dialetheist will have to face difficult challenges.

3.3. Problems for the dialetheist's modeling

So, we have seen that there is a clear difference between the traditional formal approach to the Liar and the dialetheist's account of it. In order to allow that some sentences be true and false, the Exclusion principle was abandoned, leading us to distinguish falsity from untruth. As we have also seen, Priest doesn't think that this will cause dialetheism any trouble. Of course, there is a discomfort that the extended version of the Liar is "more contradictory" than the simple version, but how that can be harmful? Someone accepting contradictions can certainly deal with both versions of the Liar, right? Not so, we shall argue in this subsection.

The first problem we shall point to is the fact that the failure of exclusion in order to allow for gluts has as a disastrous consequence that there are two kinds of contradictions that are indeed *different*. Given a sentence $T(\ulcorner\lambda\urcorner)$, we may reasonably ask

for its contradictory pair. As Priest (2006b, p.78) is eager to grant, contradictories "are unique — at least up to logical equivalence"! That is, there is *only one* contradictory corresponding to $T(\ulcorner\lambda\urcorner)$.[14] Which is it? The internal $T(\ulcorner\neg\lambda\urcorner)$, or the external $\neg T(\ulcorner\lambda\urcorner)$? They are not equivalent, recall: one cannot go from an internal contradiction to an external contradiction (so, they are not logically equivalent). What results is that there are two distinct semantic principles of non-contradiction:

LNCI: No sentence is true and false at the same time.

LNCE: No sentence is true and not-true at the same time.

This fact presents challenges for the dialetheists' account of the Liar. It seems that the reasonable conclusion is that at least one version of the Liar will not be a genuine contradiction under the dialetheists' treatment! There is a constant equivocation between the two versions of the law and their respective versions of contradictions, but it is important to notice that they are no longer equivalent in the dialetheist treatment under consideration here. Let us explore the two possibilities, *viz.*, that the contradiction envisaged by dialetheists is of the kind forbidden by LNCI, and also, that it is of the kind forbidden by LNCE. We begin with the first one, that is, we begin by discussing the case where one assumes that a contradiction is a sentence that is both true and false.

Internal contradictions

It seems fair to say that by banning exclusion the dialetheists wish to claim that being true and false is a contradiction, and also a glut. That view also seems to follow from some definitions of dialetheism, as mentioned in section 1, when it is defined as holding that some sentences are true and false, so that the "true contradictions" that it accepts are of the internal kind. This approach has the advantage of turning true contradictions into gluts. Also, the formal treatment in LP explicitly requires this version: the paradoxical sentences in this case are the true and false ones. So, as an option, this one seems to have textual evidences that privilege it over the alternative proposing to ban LNCE and allow for external contradictions. If this is correct, the idea seems to be that the contradictory of a sentence is its contradictory in the internal sense. This option is problematic for the Liar, though.

The first trouble for this option for 'contradiction' is that when contradictions are understood as internal contradictions, the extended Liar ends up not represented by the present account. That happens simply because the extended Liar is not presenting a contradiction like that, but a rather stronger case of opposition. The dialetheists face difficult questions here, given that both paradoxes were intended to represent 'the same old problem'. In fact, given that the extended version of the Liar is not

an internal contradiction, it seems to fail in characterizing what is involved in a legitimate version of the Liar: the switch between *Bona fide* truths and the Rest. The extended version in this case switches between truth an untruth, not between truth and falsity, which would be the legitimate contradiction once we assume, as we are for the sake of argument, that a contradiction is an internal contradiction (and recall: falsity and untruth are not equivalent for the dialetheist). Given that there is no (internal) contradiction in the extended Liar, there is no legitimate switch, it seems, and we have no Liar, just as in the case of the gap approach to the simple Liar. So, the same charge that the dialetheist made on the gap approach can now be directed at the dialetheist: it fails in representing a legitimate Liar. Worst: it is the extended Liar that fails to be the correct kind of contradiction in this case!

Furthermore, it seems that if one is taking gluts to be the answer to the Liar, then only the simple Liar will have an immediate answer, given that only the SLS leads us to a glut. But then, the revenge movement may work also against this approach! How can that be? Well, allow truth and gluts to be the new set of *Bona fide* truths, — as indeed is the case in LP — and the Rest is the set containing just the untrue sentences. In other words, the *Bona fide* truths are the true sentences, along with some false ones too (the gluts). The Rest is the set of sentences that are just false. The extended Liar sentence, while claiming to be untrue, actually says of itself that it is just false, and we have the conclusion that it is true and untrue. But that is far more than true and false. True and untrue are exclusive, so that we have a situation that was not envisaged by the dialetheist's solution, who was only accounting for true, false, and true and false sentences. When we are considering only internal contradictions, there is simply no story about what happens when a sentence is true and untrue.

It could be objected that Priest, in a quote we mentioned before, had already argued that truth and untruth are not exclusive. So, there is no problem in having true and untrue sentences together, they are just more contradictory. But that does not seem to be the case, there is good evidence that Priest also thought truth and untruth to be exclusive, against his own previous claims! Consider the picture on the nature of truth we find in Priest (2004, p.27):
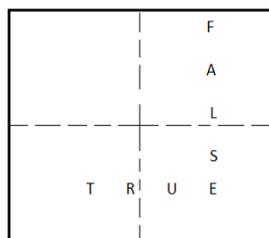
The top left corner is empty, it is the place representing gaps (exhaustion bans it

from being occupied). The bottom right corner accounts for gluts; the simple Liar is supposed to be there. Now, the sentences that are not-true are placed in the top right corner. So, truth and untruth are exclusive, if the picture correctly describe Priest's intuitions about it; there is no intersection between truth and untruth. So, how come that he could have claimed that truth and untruth are non-exclusive? Well, precisely because Priest took for granted that the extended Liar is a sound derivation of a dialetheia, a true contradiction, and so that it could also be dealt with inside the gluts approach, the standard understanding of an *internal* contradiction. However, when a contradiction is an internal contradiction, gluts and external contradictions come apart: gluts are sentences that are true and false, external contradictions are true and untrue sentences. External contradictions cannot be tamed by glut theory. In the language of LP: external contradictions are a conjunction between sentences bearing designated and non-designated truth values. In this case, the extended Liar provides for a revenge even in the case of dialetheism. The Rest class is characterized as the just false sentences, and there is a legitimate switch of truth values.

Moreover, there is a further case to be made on the separation between falsity and untruth concerning the pragmatic notions of acceptance and rejection. As is well known, if some internal contradictions are acceptable, then dialetheists argue that we may rationally accept some false sentences too; truth and falsity are mixed (let us concede it, for the sake of argument). Priest (2004, p.34) claims that the situation may be entirely different for untrue sentences: "if something is shown not to be true then one ought to reject it". So, unlike false sentences, which inherit acceptability by being mixed with true sentences, untrue sentences do not inherit the same acceptability; the reason, we may suppose, is because untrue sentences never get mixed with true ones, as the above picture shows. So, truth and untruth are exclusive.

However, the worst part of adopting only internal contradictions as our official version of allowed contradictions is that we end up with no story about how to avoid triviality in the extended Liar (we come back to this issue also in the next section). Recall that in order to have acceptable contradictions, explosion must be avoided: from $\alpha \wedge \neg\alpha$ one cannot infer an arbitrary β. The counter-model to this inference in LP comes from glut theory: pick a glutty α, so that $T(\ulcorner\alpha\urcorner)$ and $T(\ulcorner\neg\alpha\urcorner)$, and a false β, that is, $T(\ulcorner\neg\beta\urcorner)$. Remembering that in LP glut is a designated value, we have designated premises, false (and just false) conclusion; the inference is invalid. Is there any similar case for an external contradiction? There isn't. There is no story about what to do with $T(\ulcorner\alpha\urcorner)$ and $\neg T(\ulcorner\alpha\urcorner)$; recall that the opposition now is between designated value and not-designated value, an explosive combination. So, Priest is not justified in assuming that the case of external contradictions can be dealt with without providing for further argument and somehow accounting for external contradictions.

That point can be made in a different way. Consider a syntactical contradiction $\alpha \wedge \neg\alpha$. Applying the T-schema we have $T(\ulcorner\alpha\urcorner) \wedge T(\ulcorner\neg\alpha\urcorner)$. As far as the dialethe-

ists' story goes, from this internal contradiction one cannot infer an arbitrary β. That works due to the glutty character of the internal contradiction. However, by banning exclusion one cannot give a homophonic theory of truth for the syntactical contradiction, that is, one cannot go from $T(\ulcorner\alpha\urcorner) \wedge T(\ulcorner\neg\alpha\urcorner)$ to $T(\ulcorner\alpha\urcorner) \wedge \neg T(\ulcorner\alpha\urcorner)$. Then, external contradictions cannot benefit form the treatment given to internal contradictions in LP. There is no account of an external contradiction in a dialetheic theory of truth. Priest (2006b, p.85) holds that one could "endorse the claim that $\neg\alpha$ is true iff α is not true". That endorsement would allow for the homophonic theory of truth, and consequently, for ruling out explosion from $T(\ulcorner\alpha\urcorner) \wedge \neg T(\ulcorner\alpha\urcorner)$. However, it would also allow the full validity of exclusion! Recall that, according to Priest (2006a, p.67) himself, the question of whether truth and falsity are exclusive is "obviously the question of whether dialetheism is true". So, the homophonic theory of truth can be provided for at the cost of the truth of dialetheism; it is curious that Priest would concede that one could endorse the above claim accepting exclusion. The constant swing from rejecting to accepting exclusion, and back, which is occasionally found in Priest, seems to us to point to a lack of coherence in dialetheism.

The problem, to put it in a nutshell, is that even though some sentences may be true and false, still no sentence can be true and untrue, even by dialetheists' light (unless one wishes to falsify dialetheism, as Priest seems to concede sometimes). Of course, Priest would claim that truth and untruth are not exhaustive. But given the above picture and the discussion Priest makes of it, we find it incoherent to claim so.

The typical argument to ensure that truth and untruth may be dealt with as gluts, allowing for non-exhaustiveness is as follows: due to exhaustion, untruth implies falsity, so, the conclusion of the extended Liar implies a version of the conclusion of the simple Liar, which is a glut after all. In this case, the conclusion of the extended Liar can be treated by LP, right? We believe that this argument can be plausibly resisted, and here we come back to a claim we made previously, according to which the conclusion of the extended Liar does not imply the conclusion of the simple Liar; external contradictions need not be seen as implying internal contradictions.[15] Our point is rather simple. Consider again the picture above. The extended Liar grants that ELS is both true and untrue, so, as we said, it is placed both in the bottom left corner and in the top right corner. By exhaustion, a untrue sentence becomes false. The dialetheist would claim that due to this move, ELS is now both true and false, a glut. So, it would be moved, by sheer magic, to the bottom right corner, where gluts are found. Our understanding of the situation, which we find more compelling, is that exhaustion preserves the untrue sentence precisely where it is: in the top right corner. That is, it is fully compatible with exhaustion that the untrue extended Liar sentence "changes" *from untrue to just false*. This will not result in a glut, things remain exactly as they are. It would be rather strange that a sentence that is located in the just true and just false quarters suddenly changed status to become a glut. So,

while preserving exhaustion, we keep the extended Liar sentence as just true and just false. This still prevents a dialetheistic treatment of that paradox by LP; the opposition is still between designated and non-designated, while being compatible with exhaustion. There is no reason to prefer the idea that by exhaustion the conclusion of the extended Liar will become a glut, rather than our claim that it is only an external contradiction. To do so would beg the question against the non-dialetheist.

In fact, the picture helps us advancing the argument for the difference between gluts and conjunctions of true and untrue sentences. In order to allow that a sentence can be true and untrue one would need further resources, a super-glut, sentences that are true and untrue. But then, there is a further story that must be told, which the dialetheists have not yet addressed, and worst yet, if any sentence may be true, false, true and false, and also untrue, there is no longer sentences that could make a legitimate Rest class! We cannot characterize any kind of opposition, because all the classes are allowed to completely overlap. The characterization of the "essence of the Liar" is completely lost.

Ultimately, regardless of how many truth values are added to the theory (glut, super-glut, . . .), there should always be a split between the designated and non-designated truth values (or between Bona fide truths and the Rest, using Priest's terms). In order to grant the failure of explosion, the alleged super-glut should be in the designated set (Bona fide truths). But what would be its complement (Rest)? Total overlap between truth values forbids that a Rest class be created, and we miss the characterization of the essence of the paradox. If there is a genuine Rest, on the other hand, explosion threats when Bona fide truths and the Rest meet; we lose paraconsistency. This situation helps us clarify a problem with LP that will emerge more forcefully in the next section: in LP, we have that true and false sentences, taken as a glut, do not work as contradiction. By putting true sentences (including gluts) in the Bona fide truth set, we have that explosion fails and LP comes out as a paraconsistent logic. So the Rest would be just the untrue sentences (where there is no gluts). Thus, if we are taking the contradiction as a particular twisted construction which forces a sentence to go from Bona fide truths to Rest and vice and versa, an internal contradiction (glut) is not a genuine contradiction (or else the simple Liar cannot caracterize the essence of the Liar). Of course, internal contradictions keep switching inside the Bona fide truths only; it does not switch between gluts (Bona fide truths) and untruths (the Rest), as would be required for a genuine contradiction. So, even internal contradictions would fail to characterize contradictions in this scenario (more on this in the next section).

External contradictions

Now, let us briefly consider the second option, the claim that contradictions are really external contradictions. In this case, what the dialetheist is taken to hold is that

some external contradictions should be acceptable. There is one evidence that Priest could hold external contradictions as the official contradictions: recall that explaining the two kinds of contradiction, he formulated the law of excluded middle as $(Tx \lor \neg Tx)$, which, by parity of reasoning, could lead one to think that the law of non-contradiction is also formulated using external negation, *i.e.*, leading to LNCE (even in the face of the fact that Priest explicitly says that a contradiction is an internal contradiction).

But it already follows from the previous arguments that accepting external contradictions is also not a good deal. First of all, there is simply no theory about how to deal with them. The logic available to dialetheists deals with gluts, and external contradictions are a distinct kind of thing. In particular, gluts are employed to ban explosion, and it is not clear that triviality could be avoided for external contradictions. In the second place, if internal contradictions are banned as not being the correct kind of contradiction, in order to give place to external contradictions, we lose the identification of contradictions with gluts, so that dialetheism is no longer the view that some sentences are true and false (of course, this is a result from the above discussion on the acceptance of exclusion and falsity of dialetheism). This is in complete disagreement with the claims by dialetheists themselves. As a third problem, under this approach the simple Liar will not be a contradiction. Banning exclusion one cannot go from internal contradictions to external contradictions, so that it is not really a Liar.

The upshot of this subsection seems to be the following one: the dialetheist may choose as the official understanding of contradictions one of the two options, internal or external contradictions. They are not equivalent, recall. On the one hand, by accepting internal contradictions as the real contradictions, the dialetheist will have to face a revenge movement; all that is required is that we take 'untrue' as the Rest class. There is nothing to prevent such a revenge. In other words, that means that when we have to face external contradictions, as in the case of the extended Liar, there is nothing to accommodate and keep the contradiction from exploding. Explosion of external contradictions still obtains. But that is precisely what dialetheism was expected to do: to tame contradictions. Alas, it fails in doing so in the case of external contradictions, when the official understanding of contradictions means "a sentence that is true and false". On the other hand, accepting external contradictions as the official meaning of "contradiction" requires abandoning gluts, which are, so far, the only tools allowing for a paraconsistent logical treatment available for contradictions; so that again, following this path, nothing prevents explosion. Worst than that, this path will end up giving up the simple Liar as a legitimate case of the Liar. What these considerations show is that the extended Liar cannot be accommodated in the dialetheist picture. It puts forward challenges that the logical apparatus of LP is not able to meet. However, the extended Liar is not the only problem for dialetheism.

There are further problems with the simple Liar awaiting for dialetheists like Priest, as we shall now see.

4. Dialetheism saving "the Liar" from contradiction

As we have argued, by taking into account Priest's own desiderata on a solution to the Liar, the extended Liar does not work as a motivation for dialetheism; it requires resources that go beyond the grasp of even a dialetheist, causing it some embarrassment. However, let us leave the extended Liar sentence aside for a moment and ask whether dialetheists could claim that the simple Liar still provides for good reasons for the adoption of dialetheism. As we noted above, dialetheists urge that the simple Liar teaches us that truth and falsity are nonexclusive (they "overlap"), which in turn grants that some contradictions are true and that explosion fails. In this section we argue that even the simple Liar seems not to benefit dialetheists as well; there are still important issues concerning contradiction that the dialetheist has to address. As we shall argue, the role of a paraconsistent negation in the formation of contradictions is nebulous; the dialetheist will have a hard time granting that the contradiction of the intuitive Liar is adequately modeled by LP's negation.

To put it in a nutshell, the trouble we shall point to in this section lies in the interplay between syntax and semantics. Negation, as a sentential operator, belongs to syntax; however, it must obey well-determined semantical conditions in order to generate contradictions. In Priest's dialetheism, these conditions are not met, or so we shall argue. Let us concentrate for the moment on how dialetheism is defined, in several places, using negation and syntactical resources. For the sake of argument, let us take an emblematic case:

> Dialetheism is the view that some contradictions are true: there are sentences (statements, propositions, or whatever one takes truth-bearers to be), α, such that both α and $\neg\alpha$ are true, that is, such that it is both true and false (Priest 2006b, p.1).

That is, as we have already mentioned, here Priest is taking the syntactical form of a sentence to grant it the status of a contradiction. But that is certainly not enough, as we have already seen — and Priest agrees with that. Before adding the further required ingredient, we can separate two claims in the above quote:

SyntCont Sentences of the form α and $\neg\alpha$ are contradictions.

Dia Some contradictions are true.

We contend that both **SyntCont** and **Dia** cannot live together, even by the dialetheists' own standards. The issue dwells on the very idea of contradiction, as soon

as it receives an appropriately semantical definition. In fact, there is a clear *semantical* definition of contradiction that Priest also endorses, employing the resources of the doctrine of the square of opposition:

> So if α is any statement, let $\neg\alpha$ represent its contradictory. (Contradictories, unlike contraries and sub-contraries are unique — at least up to logical equivalence.) What relationships hold between these? Traditional logic and common sense are both very clear about the most important one: *we must have at least one of the pair, but not both*. It is precisely this which distinguish contradictories from their near cousins, contraries and sub-contraries. (Priest 2006b, p.78).

We have already seen that quote before, and we have stressed the uniqueness of the contradictory of a sentence: given any sentence α, its contradictory is unique, and it is precisely that which has caused so much trouble for the dialetheic treatment of the Liar, as we argued in the previous section. Now, our focus is on the very characterization of contradiction, both as a semantic notion as well as a syntactic notion. The point is: can the dialetheist approach to $\neg$ grant that it acts as a contradictory forming operator (cfo) according to this definition? We shall argue that it cannot.

To fix the terminology, let us officially define contradiction in semantic terms (SemCont):

SemCont: α and β are contradictories precisely when at least one of them is true, but not both.

This is the definition of the square of oppositions, and it seems pretty clear that it is this definition Priest had in mind in the previous quote. The first point to make clear now is that Priest clearly and fully endorses the semantic definition. **SemCont** does not appeal to features of the syntax, in particular, it does not appeal to negation. It provides for a nice characterization of the *switching movement* between truth and falsity which contradictory forming operators must somehow adequately characterize (and which is embodied in the "essence of the Liar"). How can we grant that? In semantic terms, the answer is very simple; if $\neg$ is a contradictory forming operator, as Priest says in the above quote, "we must have at least one of the pair, but not both", or, at least one of α and $\neg\alpha$ must be true, but not both. If that is granted, then, by applying **SemCont** to assign the meaning of negation in **SyntCont**, we have that **Dia** simply will not get off the ground. That is an analytic truth, it seems, a matter of definition. But even though that conclusion is correct, let us not go so fast, because that is not how Priest sees things.

Strangely, Priest does not seem to agree with that straightforward substitution in his own words. He attempts to argue that one can have contradictions, in the sense of **SemCont**, even while adding further features to a cfo that will render a contradiction

true; in particular, allowing that truth and falsity overlap does the trick. He does that by making a *detour* through syntax. The idea is that the overlap of truth values will make **SyntCont** compatible with **Dia**. Our question will be: after the required overlap joins in, can we still grant that ¬ represents a cfo as defined by **SemCont**? In other words, can we grant that the syntactical contradiction in **SyntCont** does represent a contradiction? Our answer is that it does not. So, in order to have both **SyntCont** and **Dia**, one has to abandon the intuitive meaning of contradiction, as encapsulated in **SemCont**. Let us see.

The argument goes as follows. Following Priest, we assume that **SemCont** is correct and associate it with ¬, an operator which acts on sentences to generate the contradictory of that sentence; this interpretation allows us to derive some elementary syntactical properties of negation. With that association, Priest says that we immediately grant the validity of the *syntactical* laws of excluded middle (LEM) and non-contradiction (LNC):

LEM $\Box(\alpha \vee \neg\alpha)$

LNC $\Box\neg(\alpha \wedge \neg\alpha)$

It is straightforward that if ¬ is a contradictory forming operator, that is, if it behaves as prescribed by **SemCont**, we grant that α and $\neg\alpha$ are exhaustive (we must have at least one of the pair), they exhaust the possibilities and, as a consequence, **LEM** is granted. By **SemCont** we have that α and $\neg\alpha$ are mutually exclusive (we cannot have both of the pair) and, as a consequence, **LNC** is granted. In this case, of course, we have admitted that negation is a cfo through the semantic characterization, and the syntactical laws are granted.

But that, at least so far, is true also of classical negation. Dialetheists accept those features (see Priest 2006b, chap.4), but they *add* a further ingredient: the paradoxes of self-reference. What paradoxes like the Liar prove, according to dialetheists, is that truth and falsity overlap. In fact, given that no consistent solution to the Liar is available, we should take the conclusion of the Liar at face value.

> Here we have a set of arguments that appear to be sound, and yet which end in contradiction. *Prima facie*, then, they establish that some contradictions are true. Some of these arguments are two thousand and a half years old. Yet despite intensive attempts to say what is wrong with in a number of logical epochs, including our own, there are no adequate solutions. (Priest 2006b, p.83)

Attempts to solve these paradoxes have failed, so perhaps "trying to solve them is simply barking up the wrong tree: we should just accept them at face value, as showing that certain contradictions are true" (Priest 2006b, p.83). According to Priest,

accepting the Liar then proves that some sentences are true and false, so that some contradictions are true.

Here, notice that the meaning of "contradictions" is left hanging in the air. We still do not know what is meant by such a statement; possibly Priest has in mind the syntactical characterization **SyntCont**: a pair of statements of the form α and $\neg\alpha$. What is being said, then, is that the Liar proves both to be true. Priest seems to take it to be the same as saying that truth and falsity overlap. Some sentences may be both true and false, when he claims that the Liar allows us to take "that truth and falsity overlap: for some α we have $\alpha \wedge \neg\alpha$, and so $\Diamond(\alpha \wedge \neg\alpha)$" (Priest 2006b, p.84). Notice that according to our definition, what is happening here is that a formula having the syntactical form of a contradiction is being made true, and is now being called a contradiction. That fact, the overlap of true and false, accounts for the failure of explosion: pick an α such that it and $\neg\alpha$ are true, and any false β, and there you have it.

Does that interfere with the fact that $\neg$ is a cfo, according to **SemCont**? Not at all, according to Priest. The negation being described, as we have seen, satisfies both **LNC** and **LEM**. In fact, it is a simple matter to check that, according to the truth tables of LP, if a sentence is allowed to be true and false, it still validates those laws (they are tautologies of LP). It is not a problem that we have both $\neg(\alpha \wedge \neg\alpha)$ and $(\alpha \wedge \neg\alpha)$; we have already accepted some contradictions by accepting the Liar, so, one more contradiction should cause us no harm. By allowing a sentence to be true while its negation is also true, Priest says, negation has "surplus content".

Priest reassures us that by introducing the surplus content nothing changes in the status of negation as a cfo: differently from the case of gaps, in which **LNC** and **LEM** fail, "[g]iven the conception of negation I have just described, they do not; so negation *is* a contradictory-forming operator. It may just have surplus content as well" (Priest 2006b, p.85). The account of negation described, recall, starts with **SemCont**, which interprets **SyntCont**, validates **LEM** and **LNC**, suffers an addition of surplus content through the Liar, which invalidates explosion and makes **Dia** compatible with the truth of **SyntCont**.

Our worry, however, is that by adding the surplus content, **SemCont** no longer holds, contrarily to what Priest holds. Then, taken at face value, the claim in **SyntCont** no longer holds when "contradiction" is taken from a semantical point of view. Notice that in this last quote, what is used by Priest to grant that a negation is a cfo is just the fact that it preserves the syntactical laws **LEM** and **LNC**. That is, the restrictions on a cfo are now purely syntactic. **SemCont** is used as a ladder to get **LEM** and **LNC**, a ladder that is disposed off as soon as we have the required formulas. However, we cannot forget that it was essential in motivating half of the path in establishing **LEM** and **LNC**. We are not convinced that with surplus one can still hold to the definition of contradiction that Priest started with, **SemCont**, although one

can certainly validate the required formulas. However, what happens in the end is that the negation characterized, even though it obeys the syntactical laws required, does not model the required notion of a contradiction, but rather a weaker notion of subcontrariety, and that is where failure in accounting for the Liar lies!

A return to semantical notions, independent of negation, will provide for a clear grasp of Priest's *detour*, and show that merely requiring that some formulas be satisfied is not the same as having the relevant notion of contradiction, in the target sense, captured or modeled.

Besides the semantical notion of a contradiction captured by **SemCont**, semantically we also have two further kinds of oppositions (which Priest also discusses and rejects as not encapsulating genuine forms of negation, see Priest 2006b, chap.4):

Contrary: α and β are called contraries when both cannot be true, but both may be false.

Subcontrary: α and β are called subcontraries when both cannot be false, but both may be true.

What those two further oppositions do is to describe kinds of relations between truth values, of course. One may argue that distinct kinds of negations are associated with them. Paracomplete negation $\neg_i$ is typically associated with contraries.[16] A negation that models such opposition, of course, does not validate the corresponding syntactical version of LEM, because $\alpha \vee \neg_i \alpha$ may fail to be true, although a version of LNC $\neg_i(\alpha \wedge \neg_i \alpha)$ holds. That is, in fact, one reason for Priest not to concede that such negations are genuine negations (see Priest 2006b, chap.4); without LEM there is no genuine contradiction as required by **SemCont**. Subcontrariety, as defined, is associated with a negation sign $\neg_p$ that validades LEM ($\alpha \vee \neg_p \alpha$), although it may fail LNC, given that we may have ($\alpha \wedge \neg_p \alpha$). Again, according to Priest this is not a genuine negation, given that it fails LNC. Paraconsistent negations are typically associated with subcontrariety negations $\neg_p$, given that the failure of explosion requires that α and $\neg_p \alpha$ be true. As Slater (1995; 2007) argued, paraconsistent negation in LP is a subcontrary-forming operator.[17] In fact, surplus content seems to indicate just that! How can Priest still claim that his negation is a cfo?

The issue is already addressed above. A contradiction, in the relevant semantic sense, is an opposition that is both contrariety and subcontrariety: it involves both cases together. So, by parallel reasoning with the previous paragraph, a negation forming contradictions is certainly one that validates both syntactic versions of **LNC** and **LEM**. That is a necessary condition for a negation sign to be a contradictory forming operator. But Priest goes the other way around, taking it to be sufficient: by validating **LEM** and **LNC**, he claims that the target negation sign necessarily forms

contradictions. But this second implication just won't hold. Why? Because by having the contrariety and subcontrariety, that is, contradiction, the two overlap clauses available in the definitions of contrariety and subcontrariety, *viz.*, the ones allowing that the sentences be both false, or both true, respectively, (gaps and gluts, respectively), *are ruled out*. Truth and falsity are exhaustive and exclusive. So, by allowing truth and falsity to be exhaustive and non-exclusive due to the overlap, Priest grants subcontrariety, but not contradiction in the semantic sense. The fact that the relevant syntactical laws **LEM** and **LNC** hold does not grant that the notion being modeled is the target notion; in fact, in this case it isn't!

To recap the argument: Priest begins by conceding that **SemCont** is the correct account of contradiction. He associates it with **SyntCont** in order to have a negation sign that generates contradictions. This move allows him to grant, among other things, **LEM** and **LNC**. By adding overlap between truth and falsity, that is, by granting **Dia**, he grants that explosion fails. That move, however, weakens the semantics of the negation sign, reducing it to merely a subcontrariety-forming operator. The fact that some syntactical formulas like **LNC** and **LEM** still hold under this interpretation does not mean that **SemCont** holds. It doesn't. In the end, by allowing surplus content the negation characterized by Priest fails to account for contradictions at all. There is a shift in the meaning of the negation sign that ends up weakening the opposition it represents.

As a consequence, any attempt to claim that the Liar can be represented by a dialetheistic paraconsistent negation and still make full justice to the informal paradox fails. If we agree that intuitively the simple Liar is really a contradiction, then, the negation employed will have to characterize **SemCont**, something a paraconsistent negation cannot do, given that it characterizes merely subcontrariety. Syntactical tricks, granting that some formulas of a certain kind are available, are by no means the proof that a given opposition is being correctly described.

5. Concluding remarks

So, perhaps it is time for us to rehearse the main arguments presented throughout the paper. Our main claim is that even though the version of dialetheism examined rely heavily on the Liar paradox, it is unable to deal with such a paradox in a way that would satisfy reasonable requirements accepted even by dialetheists. To recall, the requirements are: a treatment of the paradoxes must be able to account for the informal features of such paradoxes, in particular, the fact that it generates a contradiction; second, it must avoid explosion, so that it must be paraconsistent. We have spent most of section 2 arguing that dialetheists have provided for a clear-cut characterization of the "essence of the Liar": it is a construction that divides the sentences

in two main groups, the *Bona fide* truths and the Rest. What the Liar does is to provide means for a sentence to be in both sides at the same time. That is where the contradictions lies.

What we have argued here, starting in section 3, is that in order to allow gluts, Priest separates exhaustion from exclusion. Truth and falsity are exhaustive, they account for all our possibilities. But they are not exclusive: one can have true and false sentences (the gluts). Now, that could be claimed to deal with the simple Liar sentence, but it also brings further troubles. As we have argued, following Priest, is that the notions of falsity and untruth are distinct. However, that distinction is prejudicial to dialetheists. As a first point, Priest argues that truth and untruth are "more inconsistent" than truth and falsity, ranking two distinct kinds of "contradiction": one with falsity, another with untruth. There is no definitive evidence as to which one Priest and co. choose to be "the" contradictory to a Liar sentence, although there are passages which seem to privilege internal contradictions. What we end up with is that one of the oppositions to truth (falsity $T(\ulcorner\neg\alpha\urcorner)$ or untruth $\neg T(\ulcorner\alpha\urcorner)$) is not a contradiction with a claim of truth, *viz.* $T(\ulcorner\alpha\urcorner)$, so that at least one of the versions of the Liar is not a genuine contradiction. As a consequence, the two versions of the Liar are not, under this account, "the same old problem".

Were that not bad enough, there is no clear evidence that dialetheists have a treatment for the conjunction of truth and untruth, of the extended Liar. While there is a treatment for the simple Liar through the gluts (a sentence being true and false), there is nothing said so far about truth and untruth. Given that untruth is distinct from falsity, gluts are not the correct tool. So, either there is a great piece missing in the dialetheists story about the Liar, and we need further resources, like a super-glut, or else there is genuine trouble to account for the extended Liar. In both cases, as we argued, the dialetheists leave us with no account of the extended Liar that could allow us to simply accept the argument as sound. The representation of the Liar fails in accounting for the extended Liar.

Apart from troubles with the extended Liar, the simple Liar causes a lot of difficulties in the interplay between formal and informal reasoning. As Slater (1995; 2007) had already argued, negation in LP does not generate contradictions, it merely generates subcontrariety, a much weaker opposition. We have explored Priest's response to that argument and shown how he benefits from a shift from semantics to syntax. Although Slater and Priest agree about the target notion of contradiction, Priest uses it merely in order to establish the validity of the syntactical laws of LNC and LEM. Then, by adding surplus content to negation, he violates the explosion rule, granting gluts, but still holds that the negation being characterized is a cfo. However, as we have seen, to defend the negation thus characterized from accusations of being merely subcontrariety, he appeals to the syntactical validation of LNC and LEM. As we have seen, that validation is not evidence that the relevant notion of contradic-

tion, as encapsulated in **SemCont**, is being captured. Priest benefits from employing an ambiguous standard in order to evaluate whether a negation sign generates contradictions. He begins with the semantic definition; when surplus content enters the stage, it is the syntactical criteria which are employed.

We hope that it is clear that dialetheism, as so far formulated, really cannot satisfy the demands that dialetheists themselves pose on a solution to the Liar. The acceptance of "true contradictions" provides for subtle changes in the very meaning of the terms that raise the problems dialetheists wish to account for. It seems that making a contradiction acceptable makes the very idea of contradiction different. Perhaps contradictions are still the very paradigm of something we should not accept. We could, of course, change the meaning of "contradiction" and claim that what is encapsulated in LP is really a contradiction, but that, as we have seen, would still leave the gap between the formal and the informal accounts of the Liar open, frustrating Priest's aims. Formalism and natural phenomenon follow through different paths.

References

Arenhart, J. R. B. 2015. Liberating paraconsistency from contradiction. *Logica Universalis* **9**: 523–544.

Beall, JC. 2008. Prolegomenon to future revenge. In: JC Beall (ed.) *Revenge of the liar: new essays on the paradox*, pp.1–30. Oxford: Oxford University Press.

Berto, F. 2007. *How to Sell a Contradiction. The Logic and Metaphysics of Inconsistency*. London: College Publications.

Béziau, J.-Y. 2015. Round Squares are No Contradictions (Tutorial on Negation, Contradiction, and Opposition). In: J.-Y. Béziau; M. Chakraborty; S. Dutta (eds.) *New Directions in Paraconsistent Logic*. Springer: New Delhi, pp.39–56.

Priest, G. 1979. The logic of paradox. *Journal of philosophical logic* **8**: 219–241.

———. 1995. Gaps and gluts: a reply to Parsons. *Canadian Journal of Philosophy* **25**(1): 57–66.

———. 1999. What not? A defence of dialetheic theory of negation. In: D. Gabbay; H. Wansing (eds.) *What is Negation?*, pp.101–120. Dordrecht: Kluwer.

———. 2002, 'Paraconsistent Logic'. In: D. Gabbay; F. Guenthner (eds.) *Handbook of Philosophical Logic*, pp.287–393. 2nd edition, Dordrecht: Kluwer Academic Publishers.

———. 2004. What's So Bad About Contradictions? In: G. Priest; JC Beall; B. Armour-Garb (eds.) *The Law of Non-Contradiction: New Philosophical Essays*, pp.23–38. Oxford: Clarendon Press.

———. 2006a. *In Contradiction: a study of the transconsistent*. 2nd edition. Oxford: Oxford University Press.

———. 2006b. *Doubt truth to be a liar*. Oxford: Oxford University Press.

———. 2007. Paraconsistency and Dialetheism. In: D. Gabbay; J. Woods (eds.) *Handbook of the History of Logic*. Volume 3. Amsterdam: Elsevier, pp.129–204.

Priest, G.; Berto, F. 2013. Dialetheism. In: E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. Summer 2013 Edition.
    URL = http://plato.stanford.edu/archives/sum2013/entries/dialetheism/.

Slater, B. H. 1995. Paraconsistent logics? *Journal of Philosophical Logic* **24**: 451–454.

———. 2007. Dialetheias are Mental Confusions. In: J. Beziau; W. Carnelli; D. Gabbay (eds.) *Handbook of Paraconsistency*, pp.457–466. Vol. 9. London, U.K: College Publications.

Notes

[1] We focus on Priest's dialetheism here. To avoid terminological monotony, we use liberally "dialetheists" and "dialetheism" throughout the paper, but even though most of what we say apply to dialetheists in general, in this paper those terms are to be taken as restricted to Priest's view on dialetheism.

[2] To quote Priest and Berto 2013, sect.3.2: "Overall, such paradoxes as the Liar provide some evidence for the dialetheist's claim that some contradictions are *provably* true, in the sense that they are entailed by plain facts concerning natural language and our thought processes. Extended Liar paradoxes like 'This sentence is not true' are spelt in ordinary English. Their paradoxical characteristics, as dialetheists stress, are due exactly to the intuitive features of ordinary language: unavoidable self-reference; the failure of metalinguistic hierarchies, which only produce languages that are expressively weaker than English; and the obvious presence of a truth predicate for English, 'is true', which is characterized (at least extensionally) by the Tarskian T-schema."

[3] For an overview about the revenge of the Liar, see Beall (2008).

[4] See also Priest and Berto (2013).

[5] We shall employ "not true" and "untruth" as synonymous throughout the paper.

[6] Dialetheists defend explicitly this position — that there is no new contradiction in the movement of the extended Liar —; for instance, see Priest and Berto 2013, p.13 and Priest 2007, p.173.

[7] For instance, see Priest 2006a, chap.1; 2006b, chap.4.

[8] That is, a formula of the syntactical form $\alpha \wedge \neg\alpha$.

[9] That equivalence allows us to have "the same old paradox" in both cases.

[10] Priest (2006a) offers several arguments to support exhaustion.

[11] Priest 2006a, p.116 also provides arguments against exclusion due to its unnecessary multiplications of contradictions. Although this is an important part of his campaign against exclusion, we shall not deal with it here.

[12] Recall that contradictions are formulas of the form $\alpha \wedge \neg\alpha$; so, the negation seems to be outside the scope of the truth predicate.

[13] In several papers, Priest mix up falsity and non-truth. For instance, Priest claims "Suppose we define an operator, $\neg$, such that $\neg\alpha$ is true iff α is not true, and let us say, false otherwise ..." (Priest 1999, p.110). Another example is in Priest (2007) where Priest begins with the simple Liar, with falsity, in informal context representing it with non-truth in formal context:

> At its simplest, this is the claim: this claim is *false*. If it is true, then it is false; and if it is false then it is true. Contradiction in either case. To tighten up the argument, let us write T for 'is true'. Then the liar is a truth-bearer, λ, of the

form $\langle \neg \lambda \rangle$. (The angle brackets here are some name-forming device.) Now, an almost irresistible principle concerning truth, stated first by Aristotle, is that something is true iff what it claims to be the case is in fact the case; as it is usually called now, the T-schema. For every α:

$T\langle \alpha \rangle \leftrightarrow \alpha$

In particular, $T\langle \lambda \rangle \leftrightarrow \lambda$. And given what λ is:

$T\langle \lambda \rangle \leftrightarrow \neg T\langle \lambda \rangle$

$T\langle \lambda \rangle \wedge \neg T\langle \lambda \rangle$ now follows (Priest 2007, p.172).

[14] More on the concept of contradiction and its uniqueness in the next section.

[15] Recall that, as we argued in sect. 3.2, the conclusion of the simple and of the extended Liar are not equivalent.

[16] Intuitionistic negation is an example of paracomplete negation.

[17] For further discussion on the relation between paraconsistent negation and subcontrariety, see Béziau (2015) and Arenhart (2015). A negation may be both paracomplete and paraconsistent, which is usually called "paranormal".