

Dialetheism and the Inexpressibility of Exclusion

Jonas R. Becker Arenhart and Ederson Safra Melo

Abstract

Dialetheism is the view that some contradictions are true. Allegedly, one of its major advantages is semantic completeness, that is, the promise to give a natural and uniform semantics for natural language. The argument from semantic completeness to dialetheism goes via semantic paradoxes. The semantic paradoxes — as the Liar and its strengthened forms that arise through the revenge phenomena — ultimately advance a choice between semantic completeness and consistency. According to dialetheists, like Graham Priest, we should choose semantic completeness and accept that there are some true contradictions (dialetheias). However, in this paper, we shall argue that assuming true contradictions involves leaving some important notions inexpressible, ultimately leading to a weakening of one of the main advantages of dialetheism. We add further reasons for the claim that it is not possible for a dialetheist to account for the semantic paradoxes in a uniform semantics; *i.e.*, in a complete theory able to express its own semantic notions. A kind of revenge comes out in dialetheism, considering the notion of exclusion. We shall argue that, in a very important sense, inside the most prominent dialetheist setting, it is not possible to have true contradictions and semantic completeness together. In particular, it is not possible to have both an exclusion-expressing device and also a true contradiction. We conclude that the procedure to assume true contradictions not just is powerless to accomplish this dialetheist's desideratum, but also blocks the possibility to express the notion of exclusion.

Keywords: dialetheism; semantic completeness; expressibility; exclusion.

Introduction

Dialetheism is the view that some contradictions are true, that is, that some sentences of the form α and $\neg\alpha$ are both true sometimes (see Priest [6] and [7]). *Full expressivity* is considered one of the best-known arguments in favor of dialetheism. The argument arises in natural languages via semantic paradoxes

as the Liar paradoxes: natural languages are fully expressive and satisfy all the conditions that naturally lead to a contradiction. According to dialetheists, like Graham Priest, consistent solutions to these paradoxes are worthless mainly due to the revenge phenomenon: the resources put to solve the paradoxes allow the formulation of extended or strengthened versions of paradoxes [6, chap. 1], [7, chap. 4]. The process will recur again and again; the only way to avoid the proof of a contradiction consists in either avoiding some of the expressive resources of natural language or in restricting the deductive strength of natural deductive reasoning. The latter path is not an option for dialetheists, since the phenomenon to be investigated is *natural reasoning*. The first option points out that the problem about *the inconsistency of a natural language such as English* has not been resolved. [6, chap. 1]. Thus, according to this view, the only way to keep the expressive and deductive resources of natural language intact consists in embracing dialetheism and true contradictions.

Priest, Berto and Weber [11, sec. 3.2] relate the expressive resources of natural languages and paradoxes as follows:

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 or rules amounting to the transparency of truth.

This is basically the main argument for dialetheism — the one arising from the promise of semantic completeness (see Beall [3, p. 571]). Beall [3] expresses this argument through what he calls the *first "limitative disjunction"* (FLD):

(FLD): any theory of truth for a natural language will be either expressively incomplete or else inconsistent.

In other words, if a theory of truth is expressively complete, then it is inconsistent. Beall [3] advances a 'proof' for that claim, which (unsurprisingly) is just a derivation of the Liar paradox from self-reference a truth predicate.

In this view, instead of purchasing consistency at the cost of leaving some patently expressible concepts unexpressed, it is recommended that we assume

the completeness at the cost of embracing some true contradictions. According to the dialetheist view, the price is not so high, given that expressive power is to be praised above all. That is what dialetheists claim, anyway. But one should not go so fast. Expressive power seems to have a much more expensive price: as Beall [3] has argued, when full expressive power is allowed, one is bound to get into much more trouble, not only the Liar paradoxes. As a result of natural languages' expressive power, a validity predicate may also be defined, and as a consequence, a validity Curry paradox (the so-called “v-Curry”) may be derived with the same intuitive deductive resources that led to the Liar (see also [12]). However, Curry paradoxes lead directly to triviality, which is unacceptable for all (except for trivialists, but there aren't many of them). In the face of this further difficulty, Beall [3] expresses this as a *second limitative disjunction* (SLD):

(SLD): any theory of truth for a natural language is either expressively incomplete or else trivial.

In other words, if a theory of truth is expressively complete, then it leads to triviality. The ‘proof’ is the derivation of a validity Curry paradox (again, see [3] Beall for the details and further references). As a result, it seems, full expressive power will have to be restricted. That, however, severely threatens even the possibility of motivating dialetheism by a derivation of the Liar.

So, from the fact that one may derive a version of v-Curry when full expressive power is present, Beall argues that some restrictions on expressive power will have to be made; having full expressive power leads one directly to triviality. In this paper we shall argue that one needs not go as far as deriving a validity paradox (which is controversial anyway, see [10]). We claim that the adoption of the dialetheist strategy of embracing true contradictions to accommodate the Liar paradoxes already involves expressive limitations that betray the main claim for dialetheism on granting expressive completeness. In other words, we shall argue that once some true contradictions are assumed to hold (which means that some sentences of the form α and $\neg\alpha$ are allowed to be both true sometimes, recall), one loses the ability to express some concepts that would otherwise be expressible; of particular relevance for us is the idea that some kind of exclusion holds between pairs of concepts (‘true’ and ‘false’ being the most prominent ones). Our claim is that it is not possible to have it both ways: true contradictions and full expressive power do not mix.

The structure of this paper is as follows. In section 1, we present the basic ideas motivating dialetheism in relation with the Liar paradox. The expectation that one may embrace true contradictions while retaining full expressive power is explained. In section 2, we present an argument to the effect that

some exclusion-expressing resources are not available to the dialetheist. Once true contradictions are embraced, we miss the power to express exclusion of concepts. Once full exclusion is obtained, there are no true contradictions. A choice must be made between allowing true contradictions and expressing full exclusion. In section 3, we discuss some objections to the arguments of the previous section. The idea that dialetheists cannot have full exclusion-expressing devices is already known, but, as far as we know, it has never been used to the effect that there are limits on what a dialetheist may express. We hope this discussion may bring further light to our thesis. We conclude in section 4.

1 The basics of the Liar

As the previous discussion makes clear, once we have a truth predicate available, along with a device to express self-reference, we are able to express the simple Liar sentence:

$$(\lambda) \quad \lambda \text{ is false.}$$

In order to determine the truth-value of λ , we may employ reasoning by cases, which leads us to the conclusion that λ is true and λ is false. Dialetheists claim that we should stop there, assuming the conclusion as it is. In this case, λ is said to bear a truth-value glut (whether bearing a glut, being a true contradiction and being a contradiction simpliciter are equivalent in any sense is another issue; see [1] for discussion). The relation with a contradiction is thought to be simple, at least at first sight. Let T and F being the truth and falsity predicate, respectively, and $\ulcorner \alpha \urcorner$ the name of sentence α . Considering that falsity is defined as the truth of the negation, — $F(\ulcorner \alpha \urcorner)$ means $T(\ulcorner \neg \alpha \urcorner)$ (see [6, p. 64]) —, if α is true and false we may easily infer, from properties of the truth predicate, that we have both α and $\neg \alpha$. As Priest [6, p.4, fnt.4] put: “if $\alpha \wedge \neg \alpha$ is a true contradiction, α “faces” both truth and falsity”.

Not everyone is convinced that this is the only option. A well-known approach to the problem consists in adopting truth-value gaps; *i.e.*, sentences that are neither true nor false (see [5], for the classical reference). The assumption of λ being neither true nor false (*i.e.*, bearing the gap as its truth-value) avoids the need to accept true contradictions — at least for the time being.

The problem with this approach is that it is an easy prey to revenge paradoxes. One can always introduce a new paradox in terms of the solution that was given to the paradox; one can simply formulate an extended version of the paradox by following a strengthened Liar sentence:

$$(\sigma) \quad \sigma \text{ is untrue.}$$

Being untrue means that the sentence is either a gap or else false. In any case, if σ bears a gap, then, what it says is the case, and it is true. If σ is false, then, again, what it says is the case, and σ is true. So, if σ is untrue, then it is true. On the other hand, if σ is true, then what it says is the case, and then it is untrue.

We now have the same phenomenon that appeared in the case of the simple Liar, when we were confined to only two truth-values, but now with three truth-values appearing. The new Liar sentence keeps switching its truth-value among the true and the untrue. According to Priest [6], this phenomenon is inevitable in any kind of solution, and it reveals the significance of extended paradoxes. Priest [6, p. 23] explains it in terms of what he calls “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.

The idea is really simple: once some set of new truth-values are introduced hoping to avoid the contradiction derived by the Liar, a new strengthened paradox may be derived by re-classifying the truth-values into two broad *exclusive and exhaustive* classes of truth-values: the *bona fide truths* and the *rest*. The bona fide truths, obviously, contain those truth-values that count as true, while the rest contains those that are false (in logical terminology, this is a separation between designated and undesignated values). The strengthened Liar uses the rest just as the simple Liar used ‘false’ to generate a contradiction, a sentence that has truth-values in both classes. Ultimately, the essence of the Liar reveals the resilience of the paradox, displaying the apparent discomfort between consistency and semantic completeness. Priest [6, p. 24] argues that it reveals a classical chimera:

The paradox-solving problem is to produce a consistent theory that can express its own semantic notions. But this is a classical chimera: if a theory is to give an account of its own semantics, it must give an interpretation of some kind for the language of the theory.

In face of the essence of the Liar and conceding the primacy of semantic completeness, Priest argues that we have no other option except to accept that some contradictions are true.

Paradoxes of this kind are apparently valid arguments, often very simple arguments, starting from things that seem obviously true, but ending in explicit contradictions. Unless one can fault them,

they establish dialetheism. Though many arguments in the family are, historically, quite recent, paradoxes of the family have been known now for close to two and a half thousand years. It is a mark of their resilience that even now there is still no consensus amongst those who think that there is something wrong with them as to what this is. Better, then, to stop trying to find a fault where none exists, and accept the arguments at face value. (Priest, [8, p. 171]; the same argument may be found in Priest [7, p. 83])

Those paradoxes end up in explicit contradictions once an appropriate negation is introduced. The idea is that when the truth-values are divided into bona fide truths and the rest, an appropriate negation sign (a so-called *contradictory-forming operator*) is one that shifts truth-values between those two classes. With that, the paradox says that some sentences must shift between the two classes, that is, that the sentence and its negation must both be somehow true.

2 Revenge for dialetheists

The main desideratum for a correct representation of the Liar paradox, then, seems to require that we preserve the essence of the Liar: the Liar sentence (in any legitimate version of the Liar) should keep switching between taking values at the bona fide truths and at the rest class. The Liar sentence cannot be consistently accommodated in one of the classes: provided that it is in one of them, it ends up being in the other too, and vice versa, unless the paradox is incorrectly represented. Also, recall that the Liar and the reasoning leading to this switching of truth-values is a result of the expressive and deductive resources of natural language. The dialetheist's plan is, then, that one should keep that expressive power while embracing the resulting contradiction (see our previous quote of Priest, Berto and Weber [11], and Priest [6, chap. 1]).

As we see the situation, however, things are not that easy. The problem is that once one becomes a dialetheist, *something is lost precisely on what concerns expressive power*. As soon as one tries to correct that limitation, however, one loses dialetheism. A kind of revenge emerges, since that correction process may be attempted again and again, always with no accommodation of both features. There is no way to have it both ways. Let us check.

Our main claim starts by showing that dialetheism does no better than the gap approach to the simple Liar. It also fails, in its own terms, in accommodating the switch between bona fide truths and the rest (that is, in providing for a correct representation of the Liar). The reasoning is parallel to the one employed in the case of obtaining a strengthened Liar for the gap approach. In

the case of gaps, once three truth-values are introduced, with the Liar sentence bearing a gap, ‘false’ comes out as the only proper part of the rest class and, thus, we lose the essence of the Liar. There is no longer a switch between the bona fide truths and the rest. According to Priest [6, p. 23] “this solves the problem only at the cost of showing that it was inadequately posed”.

From this point of view, the revenge phenomenon simply reminds us that *it is useless to try to solve the problem formulating the paradox in terms of a category that does not coincide with the rest class*. So, in order to proceed to a legitimate version of the Liar, we just divide again the resulting class of truth-values $\{t, f, \text{gap}\}$ into the bona fide truths and the rest, with the first represented by $\{t\}$ and the latter by $\{\text{gap}, f\}$. What results from this is that the strengthened Liar sentence switches between these two classes, and we have a legitimate paradox again (following Priest, we say that now the problem now is “adequately posed”). The simple Liar sentence, in this scenario, fails to represent the Liar because once it receives gap as its truth-value, it no longer switches between the two classes. It no longer displays the essence of the Liar because it was not formulated in terms of a category that coincides with the rest class.

For the record, notice that the new division of the three truth-values between two classes classifies sentences according to two classes to which they may belong on what concerns their truth-values:

- (i) Those belonging to the bona fide truths, that is, sentences whose truth-value belongs to $\{t\}$
- (ii) Those belonging to the rest, that is, sentences whose truth-value belongs to $\{\text{gap}, f\}$.

Considering the essence of the Liar, the two classes are mutually exclusive and exhaustive of the possibilities. In the attempt to solve the problem with the introduction of gaps, we had an accompanying negation sign that is exclusive, but not exhaustive; that is, no sentence and its negation may be true, while some sentence and its negation may both fail to be true. This is how the Liar sentence gets accommodated. However, according to the new division between the two classes, a new kind of negation may be employed, one that switches sentences between the bona fide truths and the rest: $\neg\alpha$ is true iff α is untrue (that is, either false or gap). This must be done because a legitimate negation, according to dialetheists, must be a contradictory-forming operator, that is, it must shift sentences between the bona fide truths and the rest class. This is most relevant if one wishes that a sentence and its negation express a contradiction.

As Priest [7, p. 79] says:

A genuine contradictory-forming operator will be one that when applied to a sentence, α , covers all the cases in which α is not true. Thus, it is an operator, \neg , such that $\neg\alpha$ is true iff α is not true, i.e. is either false or neither true nor false. (In English, such an operator might be something like: it is not the case that.)

A contradictory-forming operator, then, is one that captures the very idea behind the essence of the Liar. It switches truth-values between the bona fide truths and the rest. That is what one must have in order to capture the idea that the Liar ends up in contradiction. The result now, however, is that whenever we have both α and $\neg\alpha$ we have a pair of sentences that cannot be both true: this set has no model. The result is that, under the typical definition of consequence, explosion results from the new kind of contradiction.

That would be all fine if it threatened only the gap approach. However, by a similar pattern of reasoning, the dialetheist now has a similar fate. Once we assume that the simple Liar sentence is both true and false, *i.e.*, a glut, we are able to introduce a new division in the class of truth-values, just as we did in the case of the gap approach. In fact, once some sentences are allowed to bear gluts as their truth-values, then, we have three classes of sentences, those bearing just truth, those bearing just falsity, and those that are true and false. In this case, the gluts are part of the true sentences, so that they fall on the new bona fide side. We have that the bona fide truths are $\{t, \text{glut}\}$, and the rest is $\{f\}$. Again, notice that we introduce an exclusive and exhaustive classification of all the sentences on what concerns their truth-values:

- (i) Those belonging to the bona fide truths, that is, sentences whose truth-value is in $\{t, \text{glut}\}$.
- (ii) Those belonging to the rest, that is, sentences whose truth-value belong to $\{f\}$.

In a totally analogous fashion as before, one may now introduce an appropriate negation to switch between those two classes: $\neg\alpha$ is false iff α is true or glut.

Again, with this stronger negation, which excludes cases where some sentence may bear a truth-value that is in both classes, we no longer are able to keep both α and $\neg\alpha$ as being both true (*i.e.*, both belong to the bona fide truths class). With the typical definition of validity, once α and $\neg\alpha$ are present, we have no model, and explosion follows. Now we have no resources to avoid the explosion, we lack something like a “super-glut”, able to overlap those two classes. With no model like that, triviality follows and the coherence of dialetheism is threatened.

The dialetheist suffers from the same fate as the gap theorist. While the gap theorist cannot grant that truth and falsity exhaust the possibilities, the glut theorist cannot grant that truth and falsity are exclusive. In both cases exhaustion and exclusion are granted once new sets of truth-values are formed, with bona fide truths and rest classes properly represented. An appropriate negation able to shift sentences between the two classes is introduced in both cases. Obviously, explosion results in both scenarios when α and $\neg\alpha$ are present, because we have restored exhaustion and exclusion for negation. So, when the essence of the Liar obtains, there is a contradiction, but no true contradiction; when a true contradiction obtains, truth and falsity overlap (some sentence is in both sides), but then we have no shift between the two classes, and a new appropriate division must be provided for. In this case, although exclusion and exhaustion between the truth-values is achieved, there is no possibility of having a sentence in both sides, so, no true contradiction if the appropriate negation is introduced.

A dialetheist may not be completely comfortable with that. Recall that according to the dialetheist, we must keep full expressive power (recall the primacy of semantic completeness). So, a dialetheist may argue, after the rearrangement of truth-values is performed between $\{t, \text{glut}\}$ and $\{f\}$, one may employ the newly defined negation to obtain a new paradox:

(ρ) the truth-value of ρ is in the rest class.

The problem is that this new sentence generates a contradiction in a scenario that does not tolerate contradictions (this is an appropriate formulation of the Liar, in the dialetheists' own terms, given that it is formulated with categories that coincide with the rest class). As we mentioned, a contradiction will explode the system. As a result, the dialetheist will say, one should take the derivation as it is and accept that in the new scenario, some sentences may be both in the bona fide truth class and in the rest class. That is, there is something, like a super-glut, that allows the overlap between those classes. That proves that there are some true contradictions again, problem solved, no?

Notice that this means leaving exclusion of truth-values out again. The two classes are no longer exclusive, and the negation in this new scenario is not properly a contradictory-forming operator. Without a clear separation between bona fide truths and rest, the essence of the Liar is lost. However, to fix things up one could define once again new classes of bona fide truths and the rest, now taking into account the new contradictions obtained with ρ , define a new exclusion expressing negation and generate a new paradox leading to a legitimate contradiction. But, again, in this new scenario, it is necessary to extend the designated class (the bona fide truths) with a new resource — something like a super-super-glut — to avoid the triviality that results from

a contradiction. That could go on forever, with exclusion between bona fide truths and rest leading to explosion, and being fixed by an overlap, of the two classes, just to lose exclusion and the essence of the Liar. Ultimately, regardless of how many times the bona fide class is extended, adding new resources to avoid the triviality, there should always be another exclusive split between bona fide truths and rest in order to preserve the essence of the liar (*i.e.*, the relevant meaning of contradiction in the dialetheist view). However, with this exclusive split, explosion and triviality appear again.

What is the lesson to be learnt from that? We think that the lesson is clear: an exclusion expressing device, a negation that grants exclusion and forms contradictions, is incompatible with true contradictions. One cannot have a negation that both switches truth-values between excluding classes of truth-values and also tolerates that some sentences be in both classes, so that there is overlap between them (that is, they are not really excluding). In this sense, there is a limitation to what a dialetheist may express: she can never express that some concepts exclude each other. That is not new (it has been investigated by Berto 2014, for instance), but it has never been used to show that one of the main tenets of dialetheism, expressive completeness and the acceptance of contradictions resulting from the Liar, *does not hold for dialetheism*. Considering that, it seems the dialetheist will have to face a difficult new and simpler third *limitative disjunction* (TLD):

(TLD): any theory of truth for a natural language is either semantic complete or allows for true contradictions.

3 Exclusion lost and found

Let us recap what has been achieved in the previous section. We have seen that dialetheists have two general claims, one about the need to keep the expressive power of language intact in any solution to the paradoxes, and another to the effect that the paradoxes should be accepted as a proof that some contradictions are true. We have argued that one of the main reasons why the gap approach to paradoxes is misconceived, according to dialetheists, is also a reason for us to see dialetheism as a misconceived approach to paradoxes. It also falls prey to a kind of revenge paradox, by the same kind of reasoning that leads to a revenge paradox in the case of the gap approach. However, the dialetheist always has the option of embracing one more contradiction as true. As we have shown, that means that the very idea of some truth-values excluding others is lost; that is achieved by weakening the properties of negation, so that it no longer switches truth-values between bona fide truths and the rest. There is a

kind of loop that starts here. Once such a weaker negation is introduced, one may always produce a new separation of truth-values and get a contradictory forming negation (one that grants exclusion and exhaustion of truth-values). But then, due to self-reference, a new paradox arises, the essence of the Liar is re-gained. Dialetheists recommend that we accept the new contradiction as true, and it all starts again.

We have argued that what this proves is that one cannot have both an exclusion-expressing device in the form of a negation and also a true contradiction. In this section we shall check some arguments that a dialetheist could advance to prevent this conclusion from holding. We shall argue that a dialetheist cannot coherently save the desired negation while retaining the desire to have full expressive power. In the end, Beall [3] was right in claiming that perhaps full expressive power will have to go, but here we show that the conclusion holds but for different kinds of reasons, related to the very motivation for dialetheism. In this sense, we believe, the problem pointed to here is more basic.

As a first reply, a dialetheist could urge that the negation in a dialetheist logic (LP, for instance) is an exclusion-expressing device, since it captures the primitive meaning of contradiction. The initial criticism comes from the well-known Slater's paper [13], where he argues that negations in paraconsistent logics, like LP, are not contradictory-forming operators (cfo). Of course, it is difficult for dialetheists to agree with Slater's claims; given that dialetheism is the view that there are true contradictions, the meaning of contradiction plays an important role in dialetheism. According to Priest, the primitive meaning of contradiction (which agrees with the "essence of the Liar") comes from the concept of contradictoriness:

"A and B are contradictories if you must have one or the other, but you can't have both" (Priest, [9, p. 467]).

Following this line of reasoning, Priest claims that we have, from the concept of contradiction, that for any sentence α , $\alpha \vee \neg\alpha$ and $\neg(\alpha \wedge \neg\alpha)$. That is just another way to frame the very idea of contradiction. Then, Priest's answer to this problem, in 'Reply to Slater', is simple: "[s]ince LP satisfies these conditions, its negation symbol is a cfo" (Priest, [9, p. 467]).

It seems that Priest thinks these two syntactical formulations of the law of excluded middle (LEM) and law of non-contradiction (LNC) encapsulate, respectively, the exhaustion and exclusion (see also Priest [7, p.78]). If it does so, then dialetheism has an exclusion-expressing device in the formal language of LP. However, this point is disputable. Considering Priest's condition to evaluate the legitimacy of a sign of negation, $\alpha \wedge \neg\alpha$ just grants the contradiction

if the ‘ \neg ’ is a cfo. So, if we have a logic with a negation that is not cfo, the syntactical formulation $\alpha \wedge \neg\alpha$ is not granting the meaning of contradiction. Although $\alpha \wedge \neg\alpha$ has the syntactical form of a contradiction, it does not encapsulate the concept of contradiction (as contradictoriness). The meaning of contradictions is prior to theories of negation (this is clear in [7, chap. 4]).

Keeping this point of Priest in mind, we consider that the symbol of negation in LP is not a cfo; particularly, it is not an exclusion-expressing device. To try to make this delicate point clear in a simple way, consider the reason that makes LP grant both $\alpha \vee \neg\alpha$ and $\neg(\alpha \wedge \neg\alpha)$. These two formulations hold in LP because glut is being taken as a legitimate truth-value in advance; without the availability of glut, $\alpha \vee \neg\alpha$ and $\neg(\alpha \wedge \neg\alpha)$ does not hold any more (recall: the meaning of contradiction has priority). However, if we assume glut as given beforehand, we lose the primitive notion of exclusion that comes from of the previous concept of contradiction (*i.e.*, “you can’t have both”). Once glut is added, the interpretation of negation changes, and so we cannot naturally say that the meaning of exhaustion and exclusion is being granted (See [2]). Thus, the formula $\neg(\alpha \wedge \neg\alpha)$ holds in LP at the cost of the violation of primitive meaning of exclusion. So, from the fact that some formal theory holds a particular formula as valid it cannot be inferred that the primitive notion of exclusion is being formally captured. Whenever we have a true contradiction (expressed as a glut) we cannot have the meaning of exclusion. It just corroborates our point again.

As another try, a dialetheist may object to the above argument as follows. The argument was elaborated on the basis that the attribution of truth-values is consistent, that is, that one can consistently attribute truth-values to a sentence, so that some sentences, for instance, end up being true and just true, false and just false, or true and false. It is this assumption that allows us to coherently separate between the bona fide truths and the rest class. In other words, it seems we assume that the truth-value attribution is performed in a consistent metalanguage. Once one admits of inconsistent truth assignments, then, it is said, the argument above will not work. There is no possibility of separating between two classes of truth-values consistently.

Our answer to that is quite simple. Once one admits that no truth-value can be attributed consistently, one is just corroborating our criticism. If the possibility of a sentence bearing a glut infects every level of language, then exclusion is never explicitly expressible. That saves dialetheism from a revenge movement, but leaves it with limitations on the expressive power, which, recall, were one of its main alleged advantages. In other words: if there is always the possibility that a sentence is a glut, and in particular the sentence that makes the attribution of truth-values, then, one can never make sure that some sentences are just true and others just false. That is a well-known feature

of dialetheism, but for that very reason dialetheism is lacking in expressive power: it cannot be expressed that a sentence has a single truth-value (such as true, and just true). Dialetheism is saved from the revenge paradox, only to fall prey of expressive limitations. Again, there is no way of having both a contradiction that encapsulates exclusion and true contradictions (truth-value gluts) together.

The problem may be further developed as follows. Consider that α is a statement attributing a truth-value to a sentence, let us say, it represents the sentence "The sentence s is true". According to dialetheism, one cannot grant that α is just true. It may also be false, so that truth attribution is inconsistent, and the revenge movement is not warranted. Now, besides the expressive problem we have already mentioned, there is a further difficulty here. What motivates the glutty attribution to such sentences? In the case of the Liar, it is clear that the Liar sentence receives a glut because it ends up true and false after a reasoning the dialetheist accepts. What to say of the sentences making the truth-value attribution? They certainly are not the result of a deductive reasoning such as the Liar, they are adopted as gluts to save dialetheism from incoherence. But then, besides the problems of expressive power, gluts must be operating beforehand, and they are required even for dialetheism to be coherent. That is certainly very unconvincing, given that it is the very idea of glut that is in question.

Perhaps the dialetheist could claim that an exclusion expressing device is not so closely related to negation in natural languages, so that nothing is missing. There may be other ways to make others understand that we mean that this excludes that, and these other ways do not need to employ negation, precisely because negation is always compatible with the truth of both the original sentence and the negated one. For instance, by saying that John is in Paris we are explicitly saying also that he is not in Berlin, and that is done without the use of negation (although we had to explain that with the use of negation). Notice what is at stake here: this is a kind of empirical claim about what people do when they use some expressions that are typically explained with use of negation, but now they should be understood as not involving negation. However, even though one could express exclusion through such indirect means, that does not mean that people do so all the time. As far as we know, no one conducted such an investigation, and it is hard to believe that people detach negation from exclusion. When one says that Socrates is not a lemon one is not also contemplating the possibility that Socrates is a lemon, and one does that precisely by saying that Socrates is not a lemon. Dialetheism seems to fail to get that fact of natural language.

A dialetheist could, as Priest does, claim that this is going too far. True contradictions are rare, and one needs not a general expressive device to say

that something is a true contradiction, with other ones to say that a contradiction is false. We evaluate the evidence on a case by case basis, and in general when we use negation we assume that we are not dealing with a true contradiction. We just leave open the possibility that some contradictions be true, for the case of paradoxes.

However, that won't do. Whenever we think that a sentence with a negation is true, when we wish to express that in words we allow (according to the dialetheist) that the sentence be also true, thus opening the very possibility we thought was excluded. That opens up an uncomfortable gap between our beliefs and what we say, and also, make some thoughts out of the reach of language. Universal expressibility, in any theory willing to account for universal expressibility, should account for that fact. Furthermore, notice that once we use negation in a dialetheist setting, the negation of a sentence that may be true and false is just the same as the negation of a sentence that may be just false. There is nothing in the expression to distinguish the cases. That is why it seems that we are either always allowing for the possibility of a true contradiction, or else that language does not express our thoughts, precisely in the cases where we think no true contradiction may obtain. For an example of such situation, consider the sentence "Socrates is not Japanese". Someone that utters that sentence, we presume, is not willing to leave it open the case that Socrates may also be Japanese. However, in a dialetheist scenario, that is not also expressed, and the speaker may be left with such unintended meanings uttered too.

Those responses to possible objections reveal some of the weaknesses of a dialetheist account of negation as an exclusion expressing device. They (the deficiencies) infect the understanding of language advanced by the dialetheist, and certainly further difficulties may arise in this front. We, however, leave the issue here for now, given that exploring the consequences of dialetheism for a theory of meaning would take us far from our more modest goals in this paper.

4 Final remarks

By way of conclusion, let us recap and highlight, in general, some points that were advanced in this paper. As we have seen, one of the main arguments to dialetheism comes from semantic completeness. Dialetheists usually claim that Liar paradox and its strengthened forms that arise through the revenge phenomena lead us to a dilemma: semantic incompleteness or inconsistency (the FLD). So, taking into account the primacy of semantic completeness, the conclusion is that some contradictions are true and dialetheism is right. However, this argument is controversial, even among dialetheists. As we have seen, considering a parallel reasoning to the dilemma aforementioned, but using

a validity Curry paradox, Beall introduces a less palatable dilemma: semantic incompleteness or triviality (the SLD). Since dialetheists are not trivialists, semantic completeness has to go, preventing dialetheism from fulfill the promise of semantic completeness.

In this paper, we advanced new arguments that corroborate that idea that the full expressivity is something that dialetheism cannot achieve without triviality. But, differently from Beall's arguments, we put the problem without appeal to the validity Curry paradox. As was evidenced, the problem arises also in the scenario of Liar paradoxes. We argued that the Liar — that in general is taken by dialetheists to favor their view — displays that there are limits on what a dialetheist may express. Particularly, the sense of exclusion, related to the dialetheist view, cannot be expressed in such a view. The idea comes from the so-called “essence of the Liar” that reveals the precise sense of contradiction involved in dialetheism. To account for the essence of the Liar is described with the help of two exclusive and exhaustive classes of sentences, the bona fide truths and the rest, recall. The contradictory-forming operator switches truth-values between these two classes, and, no matter how these two classes are modified, the revenge restores the legitimate sense of contradiction ensuring the exhaustion and exclusion that are in the essence of the Liar.

Priest uses the essence of the Liar to argue that the gap approach fails to solve the problem of the Liar, since it does not put the problem adequately — it also reveals the semantic incompleteness, as we have seen. Through a parallel argument, we advanced a kind of revenge over dialetheists that is difficult for them to avoid, even assuming true contradictions. The point is: when the essence of the Liar is achieved, triviality gets in; and when the triviality is avoided, assuming some kind of true contradiction, the exclusion results inexpressible. The process may be repeated again and again, but the dialetheist desideratum — of having full-expressivity along with true contradictions — is never achieved. So, we advance an unpalatable dilemma for dialetheists: semantic completeness or true contradictions. The main problem here is not just that dialetheists have some inexpressible notion. It is not just any notion; it is the fact that a crucial conception involving the core of the meaning of contradiction — coming from the essence of the Liar — comes out inexpressible in dialetheism. After all, it is not possible to have, inside the same setting, true contradiction and a negation that grants the exclusion.

So, it seems that Beall was right about the dialetheist's failure to accomplish the promise of semantic completeness. But, our point goes way beyond corroborating Beall's conclusion. Our arguments put a burden that is difficult for a dialetheist to bear. It is difficult keep dialetheism motivated with the inexpressibility of exclusion. The main motivation to dialetheism — Liar paradoxes — cannot be accounted for in a dialetheist setting without inexpressibility of

exclusion, where true contradictions plays a crucial role in the theory. So, with Liar paradoxes out of the dialetheism setting, an important motivation to dialetheism gets lost. But the Liar and its strengthened forms, that arise through the revenge phenomena, are still there, reminding us that dialetheist desideratum — having full exclusion expressing along with true contradiction — cannot be accomplished.

References

- [1] Arenhart, J. R. B., and Melo, E. S. Is dialetheism self-coherent? *South American Journal of Logic* **3**(1): 87–109, 2017.
- [2] Arenhart, J. R. B., and Melo, E. S. Classical Negation Strikes Back: Why Priest’s Attack on Classical Negation Can’t Succeed. *Logica Universalis* **11**(4): 465–487, 2017.
- [3] Beall, JC. Trivializing sentences and the promise of semantic completeness. *Analysis* **75**(4): 573–584, 2015.
- [4] Berto, F. Absolute contradiction, dialetheism, and revenge. *The review of symbolic logic* **7**(2): 193–207, 2014.
- [5] Kripke, S. Outline of theory of truth. *Journal of philosophy*, **72**: 690–716, 1975.
- [6] Priest, G. *In Contradiction: a study of the transconsistent*. 2nd edition. Oxford: Oxford Un. Press. 2006.
- [7] Priest, G. *Doubt Truth to be a Liar*. Oxford: Oxford Un. Press. 2006.
- [8] Priest, G. Paraconsistency and dialetheism. In: Dov Gabbay and John Woods (eds.) *The Handbook of the History of Logic. Vol 8. The many-valued and non-monotonic turn in logic*. Amsterdam: Elsevier, pp. 129–204, 2007.
- [9] Priest, G. Reply to Slater. In: Beziau, J.-Y., Carnelli, W., Gabbay, D. (eds.) *Handbook of Paraconsistency*: 467–474. College Publications, London. (Studies in logic), 2007.
- [10] Priest, G. Old wine in (somewhat leaky) new bottles. Some comments on Beall. *Australasian Journal of Logic* **13**(15): 89–95, 2016.
- [11] Priest, G., Berto, F., and Weber, Z. Dialetheism. *The Stanford Encyclopedia of Philosophy*. (Fall 2018 Edition), Edward N. Zalta (ed.), 2018.

- [12] Murzi, J., Shapiro, L. Validity and Truth Preservation. In: T. Achourioti et. al. *Unifying the philosophy of truth* , pp.431–459. Springer, 2015.
- [13] Slater, B.H. Paraconsistent logics? *J. Philos. Log.* **24**: 451–454, 1995.

Jonas R. Becker Arenhart
Department of Philosophy
Federal University of Santa Catarina (UFSC)
Rua Eng. Agrônomo Andrei Cristian Ferreira, s/n - Trindade, CEP 88040-900, Florianópolis - SC, Brazil
E-mail: jonas.becker2@gmail.com

Ederson Safra Melo
Department of Philosophy
Federal University of Maranhão (UFMA)
Av. dos Portugueses, 1966 - Vila Bacanga, CEP 65080-8051, So Luís - MA, Brazil
E-mail: saframelo@gmail.com