

Malfunctionalism and the Liar Paradox: Bad Expressions of Good Content

WILLIAM BONDI KNOWLES

Abstract

Self-denying utterances like ‘What I am saying is false’ are paradoxical because a contradiction follows whether the utterance is true or not. This is the ancient Liar paradox. A popular strategy for solving the paradox, here christened *malfunctionalism* with prominent supporters including Kripke and Mackie, is to say that the relevant utterances malfunction and fail to say what they appear to say. But a credible account of how the failure might occur has so far proved elusive. The right words appear in grammatical order with appropriate speaker intentions, so what could possibly stand in the way of a successful statement? A recent development in the debate sheds new light on this question. It has transpired that malfunctionalism should regard the paradoxical utterances as bad expressions of good content: the propositions do exist, they just can’t be expressed by those utterances. Any malfunctioning must therefore pertain to the link between utterance and proposition. The headline claim of this paper is that a closer look at this link in the Liar case does indeed reveal an anomaly. On standard views, an utterance is connected to the proposition it expresses by way of an asymmetrical truth-dependence, such that the truth-value of the utterance is determined by whether the proposition obtains. But in the Liar case, this would require a truth-value to asymmetrically depend on itself, which is incoherent. Hence the asymmetrical truth-dependence breaks down, and malfunctionalism is vindicated.

1. Introduction

A common assessment of sentences and utterances which generate Liar paradoxes is that they malfunction and fail to say what they appear to say. This is *malfunctionalism*.¹ Consider

¹ Often described as the ‘no proposition view’ or ‘meaningless strategy’. I think it’s good to have an actual name for the position. ‘Malfunctionalism’ is also truer to the content of the view than such descriptions, as will become clear.

thus an Epimenidean Liar paradox where Epimenides says the following:

No Cretan statement is true.

Since Epimenides was himself a Cretan, if this statement is true, then it is not true. And if we imagine that every other Cretan statement is actually as untrue as Epimenides says, then if Epimenides's statement is also untrue, then what he says is really the case, so it's true. Either way, in the imagined circumstances, whether Epimenides is right or wrong, he is *both* right and wrong. Paradox.

Assuming that contradictions are in fact not true, something has gone awry, and there is no shortage of quite radical suggestions for what. Some follow Tarski (and Russell) and forsake customary univocal truth in favour of a fragmented hierarchical notion. Others restrict the validity of ordinary ways of reasoning, *e.g.*, by denying the law of the excluded middle which is needed to get from a contradiction 'either way' to paradox. Others still revise not the reasoning but the assumption that contradictions aren't true, and thus turn to dialetheism. The potential revisionary force of the problem is immense.

Malfunctionalism aims to avoid all this by denying that Epimenides's utterance amounts to a genuine statement. Each word may be individually meaningful, and they're assembled in the right grammatical way for an assertion, and yet somehow the words add up to less than the sum of their parts. The utterance isn't exactly meaningless but rather *propositionless*: no claim is made on the way things are; no condition is set for the world to meet or miss. Epimenides may of course try to express a proposition (and he may even have the right speaker intentions, more on which later), but the attempt, so says malfunctionalism, is a failure.

Is malfunctionalism just a desperate head-in-the-sand manoeuvre to avoid facing up to an inconvenient truth? The view is at least no modern invention, with historical roots strong enough for the position to have a name in medieval philosophy ('cassationism'²), and long enough to reach Ancient Greece (Beth 1966: 25). And the initial, kneejerk reaction to the paradox is commonly that something about the utterance isn't quite right (Beall 2001: fn. 1). Behind this intuition is, I believe, the strange way in which Epimenides's contradictory truth-values are derived. The paradoxical conclusion that Epimenides is both right and wrong is not derived from his describing a contradictory state of affairs, but rather from the way in which

² Goldstein (2001: 117).

the truth of the statement depends on the non-truth of the statement, and vice versa. Ordinary representational truth-values these are not.

The aim of this article is ultimately to vindicate this intuition. The paradoxical derivation of Epimenides's right and wrongness is odd, and it is at odds with his expressing what he appears to. The basic idea, novel as far as I know, is that Epimenides's utterance incurs a specific kind of connection failure between truth-value and intended subject matter. The argument consists of two premises: the first is the standard idea that expressing a proposition requires an asymmetric truth-dependence between the utterance and the proposition, such that the truth-value of the utterance is a result of the truth-value of the proposition; the second is that this asymmetric truth-dependence cannot take place in the Epimenides case, since it would require an incoherent self-determination. This will be the argument of [section 5](#), and the modest hope is that it will once and for all give a universally satisfactory solution to a problem critics have held against malfunctionalism since Thomas Bradwardine's time seven centuries ago: of identifying the 'something or other' which blocks Epimenides from saying what he tries to say. Failing that, at least it seems to give those critics a tricky dilemma as to which of the two premises to reject.

To build up to this argument, the strategy will be to consider the explanatory problem in light of what malfunctionalists have found it necessary to say in response to another well-known problem, the so-called Revenge problem. The plan is to start with the response to the Revenge problem in [Section 2](#) and an important implication of it in [Section 3](#). Then in [Section 4](#) three attempted explanations of malfunctioning will be presented which I think are not successful, before returning to my own proposal in the sub-sections of [Section 5](#).

2. Revenge and response

In general, 'Revenge'-style reasoning takes a purported solution to paradox and shows how its commitments lead back into paradox. In the Epimenides case, the Revenge problem for malfunctionalism is simply that if Epimenides does fail to make a statement, then, since there are no other true Cretan statements by supposition, no Cretan statement is true. But this was what Epimenides tried and supposedly failed to say. So there's a dilemma: either malfunctionalists are roped into endorsing Epimenides's claim and landed back in paradox, or they are reduced to an involuntary

silence and cannot state a trivial consequence of their view. This general issue, or versions thereof, is commonly taken to show that malfunctionalism will ultimately need to be supplemented by some sort of hierarchical solution *à la* Tarski in order to say what needs to be said (Kripke 1975: 714; Weir 2000: 161; Rumfitt 2014: 52).

What has emerged as the standard malfunctionalist response to this dilemma is to accept (non-silently) that there are no true Cretan statements, but reject that this amounts to endorsing Epimenides's claim. *Which claim?* The whole point is that Epimenides makes no claim to embrace or renounce, so the dilemma can apparently be quashed by sticking to one's guns and repeating malfunctionalism's core commitment. This response has recently found strong support at the hands of so-called 'token relativists' like Michael Clark and Laurence Goldstein at the turn of the millennium, culminating in two prominent articles in 2004 by Jared Warren (Clark 1999; Goldstein 2001; Warren 2024a, 2004b; recent endorsements also include Whittle 2018 and Rumfitt 2014). But the point has been made by malfunctionalists from throughout the post-war period. Thus Mackie says the following (with alterations to match with my terminology and example):

The argument back from the hypothesis that Epimenides's utterance malfunctions to the conclusion that it is true is incoherent: it is convincing only because we do not hold fast to that hypothesis, but take it along with the incompatible assumption that Epimenides says what he appears to. We inconsistently combine a literal, mechanical, interpretation of its content with the critically reflective judgement that it malfunctions. (Mackie 1971b: 49)³

If Mackie is right, then the only upshot of the supposed dilemma is that *we* can say what *Epimenides* can't, but that is not necessarily a damaging consequence, the malfunctionalist will say, given that Epimenides's claim is self-referential and generates paradox while ours is not and does not.⁴ One way of making sense of the malfunctionalist gambit here is as consisting of two moves. The first is destructive: on account of its malfunctioning Epimenides's utterance is removed as a candidate for truth and falsity and as something

³ See also Whiteley 1958 and Donnellan 1957.

⁴ That *we* can say what *Epimenides* can't is sometimes known as the 'Chrysippean intuition', coined by Gupta (2001: 110).

which can be endorsed. The second is constructive: Epimenides cannot say what he intends to, but others can. The involuntary silence is not done away with entirely, but the sting is taken out of it by limiting it to certain speakers in certain contexts.

How far will this solution go? Mackie himself thinks not very far and gives a paradox ('by far the toughest version of the Liar') where he deems it inapplicable. Consider the following utterance by Epimenides (Mackie 1971b: 51):

No sentence uttered by a Cretan, standardly construed, makes a true statement.

Is this sentence, as standardly construed, true? If it is, then it's not, since it is undeniably uttered by a Cretan, namely Epimenides. If it is not true, and if we again imagine that no other sentence uttered by a Cretan is true on standard construal, then what it says is the case and so it's true. Either way, whether true or not, contradiction follows. And what's distinctive about this case vis-a-vis the previous one is that it doesn't help to say that Epimenides's utterance fails to amount to a statement, because the paradox turns merely on his having uttered the sentence. What generates paradox here is not asking whether Epimenides's *utterance* is true but whether the *sentence* he utters (standardly construed) is true. Any malfunctioning in this case thus affects us as much as Epimenides, and the 'constructive' part of the malfunctionalist gambit cannot, Mackie surmises, be carried out. His conclusion ('awkward' by his own admission) is that we are prevented from saying what needs saying, since we can neither 'endorse nor deny a sentence-token of the same type with the same reference' as the one Epimenides utters.

More recent malfunctionalists like Clark and Goldstein, however, have come up with another tool to deal with this sort of case. Rather than accept that in this example the malfunctionalist is reduced to silence, they extend the strategy one step further. Here we may not be able to use that sentence to say what needs saying, but there are other ways of, in Clark's words, 'making the point' (Clark 1999: 120). In this case simple paraphrase will do the trick, for example: 'All sentences uttered by a Cretan, standardly construed, are untrue'. This alternative way of making the point is not uttered by Epimenides and so can be true without contradiction. And if it were objected that by endorsing such paraphrases one is committed also to the veracity of the sentence uttered by Epimenides, then the malfunctionalist should once again stick to their guns and reject this argument as not taking seriously the malfunctioning of that sentence. As Smiley nicely puts it, echoing Mackie: 'The fact that the conclusion of the

inference fails to express the proposition which its form would suggest, undercuts the appeal to form on which the inference relies for its validity' (Smiley 1993: 26). How to say what needs saying thus depends on what kind of thing is malfunctioning. If it is a specific utterance, then one can produce a different token of the same sentence; if it is a sentence, then one looks for different ways of making the point.

3. Implications of the response to Revenge

This response to the Revenge problem is by no means pain free. It might, to start with, be difficult to accept that it should be within Epimenides's powers to make a sentence like Mackie's 'toughest' Liar unusable just by uttering it. Are we really so vulnerable to acts of linguistic vandalism? It has also been pointed out that the solution is another blow to the pure formality of logic: to the list of factors which can undermine formal inferences (*e.g.*, contextual changes and ambiguities) we are now asked to add malfunctioning, which some find problematic.⁵ Moreover, the question remains of whether the strategy of seeking non-paradoxical ways of making the point will work for all Liar paradoxes, and whether the malfunctionalist treatment will be suitable for all the different Liars in the first place (including *e.g.*, Yablo's infinite sequence Liar).

I do not aim here to answer these questions on behalf of malfunctionalism; indeed, for the final question I think the answer will be negative.⁶ The aim is rather to highlight an implication of the response to Revenge which is both interesting in itself and which can help to illuminate how an account of malfunctioning might go. This implication is that malfunctionalism cannot deny that there exists a proposition Liar sentences and utterances appear to express. The search for alternative ways of making the point obviously presupposes that there is a point, so it's not the purported content of Liar

⁵ Priest (1993: 43) says that it is to 'wave goodbye to the project of formal logic'.

⁶ Clark (1999: 121) and Gaifman (2000: 108) argue that Yablo's (1993) infinite sequence cases pose a special problem for the strategy of finding other ways of making the point; Weir (2000) also proposes a (more contentious) example. Knowles (2025) argues that the Clark/Gaifman argument actually shows that Yablo's paradox will be out of malfunctionalism's reach altogether.

sentences which malfunction.⁷ Rather the problem occurs when the content is put into certain words in certain contexts. It's the vessel there's something wrong with, malfunctionalism now claims, not the cargo.

Consequently the no-proposition claim is very different in the paradox case from in other areas of philosophy. A wide variety of sentences and utterances have been speculated to be propositionless in the history of philosophy, especially since the 'linguistic turn' of the early twentieth century, including *e.g.*, certain category errors ('Procrastination drinks quadruplicity'), formal contradictions, utterances with presupposition failures ('This is a fine red one' where 'this' has no referent),⁸ and many more. In no such cases is it a corollary that one just needs to find a different way of making the point. The point is rather, in all those cases, that there is no point. Not so for Liar sentences, according to this solution to Revenge.

This is interesting partly as a corrective to how malfunctionalism is described in the literature, where it is routinely presented by both adherents and detractors as analogous to supposed propositional failures in other domains.⁹ It also means that any objection to malfunctionalism which aims to establish the existence of the propositions Liar sentences appear to express, such as that of Soames (1999: 193), loses its bite.

The most important implication, however, is that it changes the complexion of the explanatory project of identifying the malfunction-causing feature of Epimenides's utterance. The question now is not just why the utterance fails to express a proposition, but specifically why it fails to express the proposition it appears to express: that no Cretan statement is true. There is nothing wrong with this piece of content, but somehow it is incompatible with being expressed by Epimenides in that circumstance. How can that be? The question is one of content incompatibility, and it may be felt that it adds to the urgency of giving an account of malfunctioning. Epimenides uses the right words in the right order and has, we may assume (especially in light of the argument made in the upcoming section) the right speaker intentions. And now it is even admitted

⁷ Gaifman (2000: 96) is one of few malfunctionalists to be explicit about this. In his terminology, failed utterances 'point to' propositions, they just don't express them.

⁸ From respectively Russell (1940), Wittgenstein (1961 [1921]) and Strawson (1950).

⁹ *E.g.*, Kripke (1975: 699), Priest (1993: fn. 13), Martin (1967).

that what he is trying to say is something that can be said. What more could possibly be needed to make a statement?

This question will now be addressed, and the answer ultimately arrived at is that the incompatibility between Epimenides's utterance and the associated proposition comes down to the fact that the truth values of the utterance cannot be taken from whether that proposition obtains, due to the problematic self-determination it would require. The argument will be explained shortly. First some pre-existing accounts of malfunctioning and their drawbacks will be surveyed, with the aim of showing why – especially in light of the solution to Revenge – a new account is needed.

4. Three attempted explanations

The most prominent pre-existing accounts of malfunctioning seek to exploit one of three supposed features of Liar sentences: derivable illogicality, blocked speaker intentions, or ungroundedness. Each of these options will now be considered in turn.

First, it has been suggested that all there is to the malfunctioning of Liar sentences is precisely that they lead to paradox and contradiction. On this view, Epimenides fails to express a proposition because if he didn't, then an illogical result could be derived, and that's all there is to say about the matter. As Prior says, 'we must just let logic teach us to where the blockages will be encountered' (Prior 1960: 32; see also Clark 2000; Smith 2006). This minimal explanation, however, is too minimal. There doesn't seem to be the right conceptual connection between leading to illogicality and failing to express a proposition. These are different things. The fact that we can derive an illogical result if the proposition is expressed may raise our confidence *that* the proposition is not expressed, but it sheds no light on *how* the proposition is not expressed. A successful account of malfunctioning needs to locate some further condition on expressing a proposition, sourced from inside the theory of propositions and expressions thereof, which Epimenides violates.

A more substantial version of the illogicality approach, amended to meet this challenge, is proposed by Goldstein (2001). He begins with the Wittgensteinian thesis that contradictory sentences do not really express a proposition on the familiar basis that they rule nothing in and nothing out; they would be untrue no matter what the world was like and thus do not describe the world as being a particular way. The Wittgensteinian thesis of course is controversial enough in itself, but the bigger problem, pointed out by Weir

(2002: 32), is that even if it is true, it doesn't help to see how Liar sentences malfunction. Neither Epimenides's utterance nor other Liar phenomena are themselves formal contradictions: 'Every Cretan statement is not true' is not of the form ' P and not- P ', nor can it be transformed into a sentence of that form via paraphrase. Granted, it might be possible to derive equivalences between some Liar sentences and formal contradictions, as Goldstein (1985) has argued, by using what he calls 'classical principles' on those Liar sentences, including T-schema-style inferences. But this cannot be why Liar sentences malfunction, because if they malfunction, then such inferences are blocked for Liar sentences. The malfunctioning of Liar sentences would therefore undermine Goldstein's suggested explanation of that malfunctioning. The suggestion fails.

The second general approach tries to understand malfunctioning as the blocking of the intentions needed for a speaker to make a statement.¹⁰ To make a statement, as opposed to merely producing noises like a parrot, an utterance needs to be accompanied by the right kind of speaker intention. Martinich has claimed, not implausibly, that the required intention at least involves 'that the audience will take the utterance as representing how things are', *i.e.*, that the audience should take the utterance as true. The next step is to say that this intention is frustrated or blocked for Liar sentences, a claim which is easiest to believe for the classic stipulated Liar ' L is not true', where ' L ' is stipulated to refer to this very sentence. L says of itself that it is not true, so the content of the sentence seems to be at odds with the speaker intention required for stating it. Martinich concludes that no one can have the necessary speaker intention 'if he utters L and knows what it means'.

The problem is that the conclusion seems false. Even for the classic L it seems *possible* to have the intention that the audience take the utterance as true. It might not be a sane intention in many circumstances, but even crazy intentions are intentions. And for other paradoxical phenomena the intention is manifestly normal. Thus if we suppose that Epimenides does not know that he is himself from Crete, then his utterance is still paradoxical, and yet there is absolutely no reason to deny that he has a fully functioning intention that the audience take his utterance as true. Even a minor modification of the main Liar example is therefore enough to wrongfoot Martinich's version of the pragmatic account. The general underlying challenge for any version of the blocked-intentions account is

¹⁰ *E.g.*, Martinich (1983), Stroll (1988), Kearns (2007).

to say why the usual speaker intentions are impossible in Liar cases, rather than merely inadvisable or bizarre. Impossibility being a high bar, it is difficult to see how this challenge can be met.

The third and most popular approach to explaining malfunctioning appeals to the ungroundedness of Liar sentences. The approach has been developed in different directions,¹¹ but the unifying idea is that getting trapped in endless dependencies on further truth sentences (either through circles or infinite chains) is incompatible with expressing a proposition. On one definition, a truth-sentence ‘*S* is true’ is ungrounded and propositionless if it fails to depend on a non-truth-related world state of affairs. This is quite a wide conception of ungroundedness, since it would include sentences which attribute truth to non-semantic objects. Thus a sentence like ‘Obama is true’ would be judged ungrounded and propositionless (dubiously: it isn’t the case that among the true things in the world one finds Obama, so this sentence seems plain false). A narrower conception of ungroundedness is what Warren (2024b: 51) calls ‘non-well-foundedness’, on which ungroundedness is nothing other than the aforementioned inescapability of dependencies on further truth sentences, either because there is an endless sequence of truth sentences or because they loop back on themselves. ‘Obama is true’ is not ungrounded on this definition. Epimenides’s utterance is ungrounded on either definition, of course, because when we unpack it by looking at each Cretan statement and asking whether it is true, this will lead back to Epimenides’s utterance itself, which again unpacks into itself and so on.

The ungroundedness explanation doesn’t fall at the first hurdle like the previous two suggestions, because the feature it appeals to is indeed a feature of Epimenides’s utterance. In fact, one main advantage of ungroundedness accounts is that it looks like Liar phenomena generally will be ungrounded, both self-referential cases and Yablo-style infinite sequence Liars. Another good thing is that there seem to be a couple of possible conceptual routes from ungroundedness to propositionlessness, or at least the beginnings of such routes, both mentioned by Mackie, each corresponding to the wider and narrower conception of ungroundedness as explained above. For the

¹¹ In addition to the distinction introduced below, ungroundedness accounts come in two completely different styles. On one side are the formal approaches associated with Kripke (1975) – *e.g.*, Leitgeb (2005); on the other are informal accounts primarily from the British philosophy scene around Ryle and Mackie, see Mackie (1971a; 1971b).

wider, Mackie says (in line with certain conceptions of disquotational theories of truth) that the predicate 'true' (unlike 'man', say) requires unpacking to count as 'something definite' (Mackie 1971a: 10; see also Beall 2001). A truth sentence will only yield a proposition when the truth predicate can be eliminated, that is, when one can unpack it into a non-truth-involving sentence. For the narrower conception of ungroundedness ('non-well-foundedness'), Mackie (1971b: 45) says that 'there is nothing that it would be for this non-terminating sentence to be true', the idea being, I think, that since there is always something left to unpack for an ungrounded sentence, one always has an unanswered question as to which part of the world will bear on the truth or falsity of the sentence (what Mackie calls an 'emptiness'), which in turn is meant to be incompatible with the expression of a proposition.

It is thus easy to see why so many malfunctionalists have tried get ungroundedness accounts up and running. Unfortunately, as well as covering all Liar sentences, ungroundedness (on either construal) also covers sentences which are currently not regarded as malfunctioning, including all the alternative ways of saying what the paradoxical items fail to say. Because if we non-Cretans say 'No Cretan statement is true' at the same time as Epimenides, then any attempt to unpack our statement will immediately encounter Epimenides's utterance and thus get caught up in the endless unpacking of that utterance. It would be useless to protest that this would not occur on the grounds that Epimenides's utterance is ungrounded when taken on face-value and thus malfunctions, because when Epimenides's utterance is taken on face-value, our utterance is also ungrounded and so, by parity, must also malfunction. On the definitions of ungroundedness given, therefore, the appeal to ungroundedness is inconsistent with the earlier solution of Revenge and contrary to the whole point of this form of malfunctionalism.

Can the problem be fixed? It might seem feasible to regard Liar sentences as 'originally' ungrounded and their paraphrases as only 'derivatively' ungrounded, where the distinction could be understood in terms of which sentences are essential to the original ungrounded circle or sequence. But this fix would put an awkward constraint on the account of why being ungrounded should lead to propositionlessness, which would have to apply only to 'original' ungroundedness. The obvious worry is that meeting this constraint would effectively amount to providing a new theory of malfunctioning. Certainly the two options for clarifying the connection between ungroundedness and propositionlessness mentioned above do not solve the problem.

There is a deeper and perhaps connected tension between ungroundedness approaches and the current conception of malfunctionism. What the malfunctionist needs is to identify a *problem of expression*; they need to point to something about Epimenides's situation which stops him from stating a perfectly valid piece of content. But the impossibility of unpacking a truth sentence into anything other than a truth sentence doesn't show that there is anything unusual or defective in the relationship between bearer and content, *i.e.*, between Epimenides's utterance 'No Cretan statement is true' and the proposition that no Cretan statement is true. Ungroundedness itself implies nothing about this relationship at all, good or bad. Consider for instance an infinite chain of sentences where each says of the next that it's true – *S1*: *S2* is true, *S2*: *S3* is true, and so on endlessly. Each sentence in this chain is ungrounded, but a critic would be well within their rights to ask why the fact that the chain continues into infinity should have any bearing on the relationship between, say, *S1* and the question it pronounces on, *i.e.*, whether *S2* is true. Of course how the chain continues is relevant for how that question may resolve, but it doesn't seem to have anything to do with the question's relationship to *S1*. So the critic could just insist that the regress stops with whether *S2* is true. The truth of *S1* depends on whether *S2* is true, and that's it; that's enough for *S1* to pronounce on that issue. What *S2* in turn depends on – as long as it doesn't interfere with this first dependence – is simply not of any significance to what *S1* says.

The availability of this kind of criticism suggests that the best the ungroundedness approach can hope for is to explain what it takes for a 'truth proposition' to count as a genuine proposition; why Epimenides's truth attribution doesn't reach the status of statement. I'm not saying that all ungroundedness theorists are unaware of or unhappy with this consequence – Mackie (1971b: 41) says that Liar sentences do not 'raise a substantial issue' – but it is inconsistent with the response to the Revenge problem. According to malfunctionism as currently conceived, there is nothing insubstantial about the issues Liar sentences attempt to raise, it is just that those attempts are unsuccessful.

Ungroundedness accounts are not completely on the wrong track. The idea that malfunctioning is down to some defect with what the truth of Epimenides's utterance depends on is, I think, right. But it is the prospective self-dependence of Epimenides's utterance – that whether it is true apparently comes down to whether it itself is true – which must be the heart of this defect. This is how to ensure that other, non-self-dependent attempts to make the point won't

malfunction. Using the self-dependence of Epimenides's utterance to this effect has been suggested before, *e.g.*, by Gaifman (2000), who only regards as malfunctioning those utterances and sentences 'inside the loop', and by Skyrms (1984), who thinks that the self-reference induces an intensional context where co-extensional expressions aren't interchangeable. What is prominently missing or under-developed in these accounts, however, is the philosophical explanatory aspect. Why should the fact that Epimenides's utterance comes to depend on itself if we suppose that he says that no Cretan statements are true, mean that the supposition cannot be sustained? These writers, like so many malfunctionalists, focus on issues related to the formal semantics of a «gappy» solution to the Liar, apparently in the belief that if we take care of the logic, the malfunctioning will take care of itself. This is precisely the opposite of the methodology employed here. As Mackie (1971a: 89) puts the point: the technician 'is primarily concerned to exclude the paradoxes, to ensure at least that none of the known ones will arise within his system ... The general philosopher wants to show that they are only apparent antinomies, that the issues about which we are which we are tempted into formal contradictions are insubstantial'. Mackie's approach, like mine, is the latter, and so it's time, finally, to explain why the prospective self-dependence of Epimenides's utterance leads to malfunctioning.

5. Epimenides and failed determination of truth by subject matter

Epimenides uses the right words in the right grammatical order, has the right intentions, and what he wants to say is the sort of thing which can be said. What more could possibly be needed for the attempted statement to be a success? A thumbnail version of the argument of this section is as follows. A relatively uncontroversial requirement for sentence *S* to say that *P* is that there be an appropriate evaluative connection between *S* and whether *P*, such that the truth-value of *S* is determined by whether *P*. That is where a sentence's truth-value must come from: the state of its subject matter. Epimenides only succeeds in stating that no Cretan statement is true, consequently, if the truth of his statement is the outcome of whether there is a true Cretan statement. But this does not occur: the truth-value of Epimenides's utterance cannot be determined by the relevant facts in this case, because the relevant facts turn out to be whether Epimenides's utterance is true, so the truth-value would

have to be determined, absurdly, by that very truth-value. The utterance gets in the way of itself. That is not to say that a truth-value for the utterance cannot be derived (notoriously it can) but its contradictoriness is not inherited from a contradictoriness in the facts supposedly described but rather has a different source. And that is incompatible with the expression of that proposition by that utterance. Epimenides's attempt to say that every Cretan statement is not true is a failure.

So I will be arguing from here on in. The argument can be split into two premises, which will be defended separately. Premise 1 is that in general, for *S* to say that *P*, the truth-value of *S* needs to be a result of whether *P*. Premise 2 is that Epimenides's utterance fails to take its truth-value from whether there are true Cretan statements, and – equally importantly – that other attempts to express the proposition need not suffer the same fate.

5.1. Premise 1

A distinctive feature of truth-values (belonging to sentences, utterances, beliefs, what have you) is that it is generally possible to trace them back to two factors: how the subject matter is represented as being, and what the subject matter is actually like. Thus we can ask what the sentence 'Snow is white' is true *in virtue of*, or what *makes* it true or what it *owes* its truth to, and find that it is that it says that snow is white and that snow is indeed white. Or one can ask what makes 'The Moon is made of cheese' untrue, the answer being that the sentence says that the moon is made of cheese and that the moon is not made of cheese. Only in the special case where no claim is made at all will the truth-value fail to come down to both the semantic and the worldly fact. The requisite match won't occur whatever the facts are, because nothing is provided for them to match with.

When a proposition is expressed, the two-factor explanatory pattern is striking in its generality. It doesn't just hold when the issue at hand is plausibly independent of us and our descriptive activities, such as for the sentence 'Snow is white'. It also holds for sentences where the relevant state of affairs is under our stipulative control: the truth of "'Snow" means snow', for example, is a result of the fact that we do use the word 'snow' in that way. And the explanatory pattern also holds when the truth-making state of affairs is brought about by the relevant utterance. When a father tells his daughter she's going to school tomorrow (as a statement, not a command), this may (partially) cause her to go to school the next day. And yet, even though

the state of affairs would not have occurred had the statement not been made, the explanation of why the statement is true is still that it says that she will go to school and that she will go to school.

It would be incredible if this explanatory pattern were a coincidence, and on standard views it isn't. On the usual representationalist picture, the role of expressing a proposition (saying that things are a certain way) is to pair up bits of language with bits of the world, in such a way that the truth-value of the bit of language is determined by whether or not the relevant bit of the world is in that condition, and *just* by that: the one and only relevant question for the truth-value of a sentence *S* saying that *P* is whether or not *P*. The truth-value of the sentence must come down to whether the proposition obtains.

That there is a close logical relationship between the truth value of a sentence and the obtaining of the associated proposition needs no argument, but the reader may want to hear more about why the relationship should be asymmetric in the way described, with the truth-value of *S* a *result of* (or *traceable to*, or *sourced from*, etc.) whether *P*. It is common to define truth just as being language-world agreement, so that something is true, by definition, iff it represents things as being a certain way and things are that way. But all that immediately follows is that, by definition, when *S* says that *P*, *S* is true iff *P*. Why the extra assumption that whether *P* is the source of the truth-value of *S*?

One motivation for the asymmetry is that, as is well known,¹² the close logical relationship imposed by *S* saying that *P* cannot merely be logical equivalence, *i.e.*, that it is necessary that *S* is true iff *P*. Sentences about non-contingent matters will have many necessary biconditionals of this kind (for example it is easy to verify that it is necessary that '2+2=4' is true iff 2+2=4 and 3+3=6). There must therefore be something else to the close logical relationship between the truth of '2+2=4' and the fact that 2+2 really is 4, which is where the asymmetry comes in: that 2+2 is 4 'makes' the sentence '2+2=4' true, as it is customary to say, which really just means that the sentence's truth-value is a result specifically of that fact. More positively, the asymmetry does seem essential to expressing a proposition. When we make a claim there is always a background of facts (which may be semantical facts) against which to determine the correctness or truth or accuracy of that claim, and if someone else were to make the same claim it would have its truth-value determined

¹² See *e.g.*, Davies (1981: 28, 39).

in an identical way, and if someone else were to make the contrary claim, then who was right and who was wrong would come down to what things were actually like. This captures what is practically important about truth: that it is determined against the objective standard of correctness of reality. Moreover, the asymmetry doesn't seem ontologically mysterious. After all, having a certain subject matter is itself an asymmetrical notion; 'aboutness' is unidirectional, it goes from the describer to the described but not back again. That claims are about the facts – 'aim' at them in some sense – would seem to be ample explanation for why the facts form the standard of correctness against which truth or falsity is determined.

The representational picture of truth sketched is, I hope, general enough to be acceptable to a large proportion of philosophers, and the final part of its defence comes precisely from an appeal to philosophical orthodoxy. There may be controversies around how to fill in the details,¹³ and no doubt the case I have made for it can be improved, but the idea itself – that the truth of anything saying that *P* must come down to whether *P* – is not very controversial. In the literature it goes under the name *truth asymmetry*, is traceable to Aristotle,¹⁴ and belief in it is not far from universal. One illustration of its sway is that it is often taken as an adequacy condition on theories of truth.¹⁵ Another illustration of its sway is that even analytic truth, traditionally conceived as the sort of truth which holds by virtue of meaning alone, is these days re-worked to accord with it, or else simply dismissed as incoherent. Thus one regularly hears influential authors in the analyticity debate claim *e.g.*, that 'once a sentence has been associated with a determinate truth-condition, whether or not that condition is met ... must be left for the world to decide', or that what is responsible for generating the truth of a given class of statements is 'something bland like "the world" or "the facts"' (Hale 2013: 127; Boghossian 1996: 365).¹⁶ Given the point

¹³ The most pressing question is perhaps how to make sense of the determination of truth by fact, and whether it can be aligned to a more general phenomenon like grounding or causation.

¹⁴ *Categories* (14b, 14–22).

¹⁵ See Melkonian–Altshuler (2024) for references (especially Liggins 2016).

¹⁶ Even those not normally associated with representationalist theories of truth sometimes slip into this way of talking, *e.g.*, Quine (1963: §3) says that 'true sentences generally depend for their truth on the traits of their language in addition to the traits of the subject matter'.

of the upcoming sub-section, this makes rather surprising allies of malfunctionalists and critics of traditional analyticity.

5.2. *Premise 2*

If Epimenides's utterance complied with truth asymmetry, then the contradictory truth-value one can derive for it when taking it at face value should be traceable to the contradictoriness of its subject matter, to its both being and not being as described. But it is not so traceable. Epimenides's utterance appears to say that no Cretan statement is true, but is itself, at face value, a Cretan statement. And since there are no other true Cretan statements by supposition, the truth-value of Epimenides's utterance will come down, well, to the truth-value of Epimenides's utterance. Trying to locate the basis of the truth-value in the subject matter just takes us straight back to the thing we were trying to locate the basis of. In order to see the utterance's subject matter as contradictory, one already needs to assume the contradictoriness of the utterance's truth-values. Hence the contradictory truth-values are not a result of having contradictory subject matter.

This is not to deny that – on face value – Epimenides's utterance does have contradictory subject matter. If the utterance is both true and false, then it both is and isn't the case that no Cretan statement is true, since that is what the utterance (on face value) says. But evidently the contradictoriness of the subject matter is not the source of the contradictoriness of the truth-value; it is not what it owes its contradictory truth-value to. It couldn't be, because whether there are any true Cretan statements depends, in the specified context, on whether Epimenides's utterance is true. The contradictoriness would have to be a result of that very same contradictoriness, which is simply not possible. Nothing explains itself;¹⁷ at most there may be things which have no explanation, no origin, but that is in any case not the nature of truth-values.

The root problem is that the truth-value of Epimenides's utterance is set up to be a result of itself; that whether it is true is supposed to be determined, per impossibile, by whether it is true. That's why Epimenides cannot say what he intends to. The argument I'm proposing is in effect a *reductio ad absurdum*. One tries to sustain the

¹⁷ In the relevant sense. A caught wrongdoer asked to explain himself is not of course expected to contravene this principle.

initial assumption that the truth-value of Epimenides's utterance is determined by whether there are any true Cretan statements. But then this turns out to require that the truth-value be the result of itself. This being incoherent, the initial assumption cannot be sustained. The utterance will not take the content intended for it. It just spits it out when Epimenides tries to push it in, like a library card in a cash machine.

It might be wondered how the claim that the truth value of Epimenides's utterance isn't determined by the relevant proposition is consistent with the epistemological fact that we can reach a verdict about its truth-value (albeit a contradictory one) when assuming that it expresses that proposition. The answer is that the contradictory truth-value is not reached by reading it off from a contradictory state of affairs. If we try to decide the truth-value of Epimenides's utterance by checking whether there are any true Cretan statements, this leads nowhere at all, neither to contradiction nor to indeterminacy, but just a never-ending circle, infinitely alternating between the question of what is the truth-value of Epimenides's utterance and whether there are any true Cretan statements. To reach the contradiction, rather, one needs to take a step back and consider *that* this circle occurs, and especially the distinctive way in which Epimenides's utterance depends on itself, *i.e.*, that its truth depends on its not being true. It is this which enables the derivation of the paradoxical biconditional that it is true iff not true. In other words, rather than the truth-value of Epimenides's utterance being decided by its supposed subject matter (which would be impossible for just leading to the circle), the supposed fact that the truth-value comes down to whether *P* is used to establish the self-dependence and contradiction. This is to *exploit* that supposed fact, not to *adhere* to it. Just because one can derive a truth-value for an utterance from the premise that it is determined by whether *P*, that doesn't mean the truth-value is derived from whether *P*.

The argument generalises for all 'self-dependent' paradoxes, including Epimenides's utterance, the classic stipulated liar *L* ('*L* is not true'), Liar loops where utterance *A* is that utterance *B* is untrue and *B* is that *A* is true, and Curry conditionals with untrue consequents.¹⁸ In all these cases is the truth-value supposedly determined

¹⁸ (a) A Curry conditional is a sentence of the form *A*: If *A* is true, then *B*, the paradox being that one can apparently prove that *B* is true no matter what it is, even when it is blatantly not true. When *B* is not true, however, the truth of *A* comes to depend on the truth of the antecedent, which is *A* itself, and malfunctioning occurs. Hence untruths cannot be proved after

by something which – by accident or design – itself turns out to depend on that truth-value. And this simply cannot be. No truth-value can have its source in whether *P*, if whether *P* itself depends on what that truth-value is. This would require the truth-value to determine itself, which is not possible.

The argument also extends to Truth-teller variants of the paradox. Truth-teller sentences say of themselves that they are true and are paradoxical not for leading to contradiction but because there doesn't seem to be anything which might settle their truth-value one way or the other. For example, if Epimenides had said 'All Cretan statements are true', and all other Cretan statements really were true, then whether his utterance was true would apparently come down, absurdly, to whether it was true. Like in the Liar case, one could admittedly reach a verdict by 'stepping back' and considering *that* the self-dependence occurs, thereby concluding that the truth-value will be indeterminate. But for the same reasons as before, this indeterminacy could not be a result of the utterance having indeterminate subject matter. An illuminating contrast can be drawn here with other supposedly indeterminate cases. Suppose for the sake of argument that it's indeterminate whether a tadpole at a certain advanced stage of its development is a frog. Then the claim that it's a frog has indeterminate truth-values, but this is only as a result of the prior indeterminacy in whether it is a frog. Unlike the Truth-teller, the claim's indeterminate truth-value is inherited from the state of its subject matter.

Another potentially helpful contrast is between the Epimenides case and other examples of an utterance's subject matter being influenced by the making of the utterance. Such influences are commonplace and don't usually entail malfunctioning. Suppose for example there are no other Cretan statements and that some Cretan, the Minotaur, say, says:

No Cretan statement is ever made.

The very fact that the Minotaur says this undermines the veracity of what he is saying; the making of the statement makes it false.

all. (When *B* is true, on the other hand, then the conditional can be regarded as true for having a true consequent.) (b) Note that the list does not include Yablo style infinite sequence Liars, which aren't self-dependent and which it has been argued are resistant to the gambit of seeking other ways of making the point (see fn. 6 above). This is a potential limitation of the account, but it doesn't bear directly on the assessment of the core cases. It may just be that paradoxes produced from infinite sequences need a fundamentally different solution to those produced from self-dependence.

Like the Epimenides case, the speaker facts influence the truth-value of the utterance. But here there is no issue with reading off the truth-value of the utterance from whether the proposition obtains, because the speaker-related facts affect the state of the subject matter, which in turn determines the truth-value of the utterance. What sets the Epimenides case apart is that, because of the would-be self-dependence of its truth-values, this determination cannot occur. The interference from the speaker-related facts doesn't just influence the subject matter but also the very relationship between the utterance and its subject matter.

Finally I want to highlight how the proposed explanation is compatible with the earlier response to the Revenge problem, which the ungroundedness approach, I argued, struggles to accommodate. In one respect what I am proposing is similar to the ungroundedness approach: Both kinds of explanation identify a problem with what the truth-value of Epimenides's utterance depends on. The crucial difference is the role envisaged for the (supposed) self-dependence. Ungroundedness accounts see the relevance of self-dependence as being that it leads to the truth-value being ungrounded, *i.e.*, that trying to unpack it just leads to a never-ending loop. The problem with this is that *our* attempt to say the same thing as Epimenides – that there are no true Cretan statements – is also, when everything's taken on face value, ungrounded. My alternative proposal sees the self-dependence as problematic in itself. It is incoherent that the truth-value of something should be determined by itself. Hence *Epimenides's* attempt to set up the truth-value of his utterance on whether there are true Cretan statements, which would lead to such self-determination, is stopped in its tracks. *Our* statement to the same effect, however, does not set up such self-dependence, not even when Epimenides's utterance is taken at face-value. This is why we can say what Epimenides can't.

In conclusion, Epimenides can certainly try to say that no Cretan statement is true, and we can assume for *reductio* that he succeeds. But the attempt to latch the truth-value of what he says onto whether there are true Cretan statements finds no edge; it slips off because the utterance itself, per face-value assumption, is a Cretan statement. And this is not a result of the subject matter being resistant to description, because other attempts to make the same point do not encounter the problem.

This may be a surprising result. It is natural to think that what we use our words to say is up to us; that we decide via conventions and stipulations and intentions which state of affairs the truth of

our utterance is to be judged against. And of course there's no question that these factors in most cases will suffice for the expression to occur. What has been found, however, is that there are circumstances in which the stipulations won't stick and conventions won't catch. There are objective limits to what can be said which sometimes take the matter out of our hands.

6. Conclusion

The aim of this article has been to consider the problem of how to explain malfunctioning in light of the answer to the Revenge problem, and ultimately to present a new way of identifying the 'something or other' blocking Epimenides's attempted statement, which looks directly at the logical relationship between Epimenides's utterance and whether there are any true Cretan statements. It has been argued that there is something about this relationship – the non-determination of truth by subject matter due to self-dependence – which a) on common views of truth and meaning means that Epimenides fails to say that no Cretan statement is true, b) extends to all self-dependent Liar paradoxes, and c) does not extend to the alternative ways of making the point.

References

- Beall, J.C. 2001. 'A Neglected Deflationist Approach to the Liar', *Analysis*, 61, pp. 126–129
- Beth, E.W. 1966. *The Foundations of Mathematics* (Harper and Row)
- Boghossian, P. 1996. 'Analyticity Reconsidered', *Noûs*, 30.3, pp. 360–391
- Clark, M. 1999. 'Recalcitrant Versions of the Liar Paradox', *Analysis*, 59, pp. 117–126
- Davies, M. 1981. *Meaning, Quantification, Necessity* (Routledge and Kegan Paul)
- Donnellan, K.S. 1957. 'A Note on the Liar Paradox', *The Philosophical Review*, 66.3, pp. 394–397
- Gaifman, H. 2000. 'Pointers to Propositions', in *Circularity, Definition, and Truth* ed. by A. Chapuis and A. Gupta (Indian Council of Philosophical Research), pp. 79–121
- Goldstein, L. 2001. 'Truth-bearers and the Liar – A Reply to Alan Weir', *Analysis*, 61, pp. 115–125
- Goldstein, L. 1985. 'The Paradox of the Liar: A Case of Mistaken Identity', *Analysis*, 45.1, pp. 9–13

- Gupta, A. 2001. 'Truth', in *The Blackwell Guide to Philosophical Logic* ed. by L. Goble (Blackwell), pp. 90–114
- Hale, B. 2013. *Necessary Beings: An Essay on Ontology, Modality, and the Relations Between Them* (Oxford University Press)
- Leitgeb, G. 2005. 'What Truth Depends On', *Journal of Philosophical Logic*, 34, pp. 155–192
- Liggins, D. 2016. 'Deflationism, Conceptual Explanation, and the Truth Asymmetry', *The Philosophical Quarterly*, 66.262, pp. 84–101
- Kearns, J. T. 2007. 'An Illocutionary Logical Explanation of the Liar Paradox', *History and Philosophy of Logic*, 28.1, pp. 31–66
- Knowles, W.B. 2025. 'Malfunctionalism and the Liar: Two Recalcitrant Paradoxes', *Thought*, 12.2, pp. 105–112
- Kripke, S. 1975. 'Outline of a Theory of Truth', *Journal of Philosophy*, 72.19, pp. 690–716
- Mackie, J.L. 1971a. 'What Can We Learn from the Paradoxes? Part I', *Crítica: Revista Hispanoamericana de Filosofía*, 5.13, pp. 85–108
- Mackie, J.L. 1971b. 'What Can We Learn from the Paradoxes? Part 2', *Crítica: Revista Hispanoamericana de Filosofía*, 5.14, pp. 35–54
- Martin, R.L. 1967. 'Toward a Solution to the Liar Paradox', *The Philosophical Review*, 76.3, pp. 279–311
- Martinich, A.P. 1983. 'A Pragmatic Solution to the Liar Paradox', *Philosophical Studies*, 43.1, pp. 63–67
- Melkonian-Altshuler, S. 2024. 'Truth Dependence Against Transparent Truth', *Asian Journal of Philosophy*, 3.2, pp. 1–17
- Priest, G. 1993. 'Can Contradictions be True?—II', *Proceedings of the Aristotelian Society Supplementary Volumes*, 67, pp. 37–57
- Prior, A. 1960. 'On a Family of Paradoxes', *Notre Dame Journal of Formal Logic*, 2.1, pp. 16–32
- Quine, W.V.O. 1963. 'Carnap and Logical Truth', in *The Philosophy of Rudolph Carnap* ed. by Schilpp (Open Court), pp. 351–384
- Rumfitt, I. 2014. 'Truth and Meaning', *Proceedings of the Aristotelian Society Supplementary Volumes*, 58, pp. 21–55
- Russell, B. 1940. *An Inquiry into Meaning and Truth* (New York)
- Skyrms, B. 1984 'Intensional Aspects of Semantical Self-reference', in *Recent Essays on Truth and the Liar Paradox*, ed. by R.L. Martin (Oxford University Press), pp. 119–131
- Smiley, T. 1993. 'Can Contradictions be True? – I', *Proceedings of the Aristotelian Society Supplementary Volumes*, 67, pp. 17–33
- Smith, N.J.J. 2006. 'Semantic Regularity and the Liar Paradox', *The Monist*, 89.1, pp. 178–202
- Soames, S. 1999. *Understanding Truth* (Oxford University Press)
- Strawson, P.F. 1950. 'On Referring', *Mind*, New Series, 59, pp. 320–344
- Stroll, A. 1988. 'The Liar: What Paradox?', *Argumentation*, 2, pp. 63–75
- Warren, J. 2024a. 'The Liar Paradox and "Meaningless" Revenge', *Journal of Philosophical Logic*, 53.1, pp. 49–78
- Warren, J. 2024b. 'Restricting the T-schema to Solve the Liar', *Philosophy and Phenomenological Research*, 108, pp. 238–258

Malfunctionalism and the Liar Paradox

- Weir, A. 2000. 'Token Relativism and the Liar', *Analysis*, 60, pp. 156–170
- Weir, A. 2002. 'Rejoinder to Laurence Goldstein on the Liar', *Analysis*, 62.1, pp. 26–34
- Whiteley, C.H. 1958. 'Let Epimenides Lie', *Analysis*, 19.1, pp. 23–24
- Whittle, B. 2018. 'Truth, Hierarchy, and Incoherence', in *Reflections on the Liar*, ed. by Armour-Garb (Oxford University Press), pp. 307–324
- Wittgenstein, L. 1961 [1921]. *Tractatus Logico-Philosophicus*, trans. by D.F. Pears & B.F. McGuinness (Routledge)
- Yablo, S. 1993. 'Paradox Without Self-reference', *Analysis*, 53, pp. 251–252

WILLIAM BONDI KNOWLES (william.knowles@ntu.ac.uk) teaches at Nottingham Trent University and writes mainly about paradoxes.