The signalman against the glut and gap theorists

Casper Storm Hansen

Abstract: Radical glut and gap theorists deny—in opposite ways—that the liar sentence has exactly one of the two values *true* and *not true*. I describe a scenario where a signalman finds himself in a situation analogous to the liar paradox: if he lights a fire at a certain time, that is analogous to the liar being true, and if he does not, that is analogous to the liar not being true. It is obvious that he must make exactly one of those states of affairs come about. It is argued that there are no relevant differences between the liar and the signalman's dilemma, implying that the glut and gap theorists are wrong about the former. A further point is that whether or not the liar is true/the signalman lights the fire, language/the signalman is misleading relative to the conditions under which the liar/the fire "ought" to be true/lit.

1 The liar and the signalman

Scenario 1: λ is the English sentence "λ is not true".

Claim about scenario 1: λ is either true or not true, and not both (whether it is true or not, the truth value it has is misleading: it doesn't represent the facts as they are to the language users).

This claim is controversial. It is denied by *glut theorists* or *dialetheists*, who think that λ is both true and not true, and by *gap theorists*, according to whom it is neither true nor not true.

Scenario 2: Once upon a time, in a kingdom far, far away, an old man lived on top of a mountain. He was known as "Kuvata" and he was a signalman. He had been sent to live on the mountain by the king so that he could make observations from his high vantage point and convey them to the king's court. He used a system as old as the kingdom itself to send the signals: light a fire at a specified time, or refrain from doing so. Each winter's solstice, a foot messenger painstakingly climbed up the mountain with a parchment containing instructions from the court. One year, they read as follows: "Each day in the coming year, you must look for enemies crossing the sea, and if they do—and only if they do—you must light a fire at sundown. You must also watch out for storms approaching from the west, and if one does—and only if one does—you must light a fire one sandglass after sundown." The parchment concluded with an instruction to

pass on a signal from another signalman, whose fires were within sight of Kuvata's mountain, but not of the palace: "Finally, you must keep an eye on whether the signalman on Mount Íslensku lights a fire two sandglasses after sundown, and if he does—and only if he does—you too must light a fire two sandglasses after sundown." That year, like every previous year, Kuvata lit a fire exactly when he ought to. Doing so was his sworn duty and a matter of honour to every member of the Signalmen's Guild.

When the next year's instructions arrived, they were brief: "Each day in the coming year, if the signalman on Mount Suomalainen does not light a fire at sundown—and only if he does not—you must light a fire at sundown." But Mount Suomalainen was Kuvata's own mountain!

Claim about scenario 2: Kuvata either lights the fire at sundown or he does not, and not both (whether he does it or not, his action is misleading: it doesn't represent the facts as they are to the court).

This claim is completely uncontroversial. Anyone would agree that in this scenario the signalman cannot defy the law that any state of affairs either obtains or fails to obtain.

What could possibly justify diverging opinions about the two claims? If Kuvata does not light the fire, he (in a certain sense) ought to light it, and if he does light it, he ought not to. The ought-to-light condition is satisfied if and only if the fire is not lit. The two cannot match. The signalman cannot represent the facts as they are. Similarly, if λ is not true, it (in a certain sense) ought to be true, and if it is true, it ought not to be. The ought-to-be-true condition of λ is satisfied if and only if λ is not true. The two cannot match. The sentence cannot represent the facts as they are.

The motivation for glut and gap theory comes down to just one thing: the assumption that the set of all true sentences must be a perfect representation of the facts of the world. Under this assumption, λ cannot be true only because if it were, the truth of "λ is not true" would misrepresent the fact that it is true and therefore "ought" to be not true, and vice versa for the option of λ being not true only. This assumption of the ability of language to represent perfectly therefore clashes with the principle that any state of affairs either obtains or fails to obtain, and not both (specifically, the state of affairs of λ being true). Glut and gap theorists believe that the principle of perfection of language must trump this basic metaphysical principle.

But why would language be able to do better than a signalman who is fanatical about representing the facts as they are? Language is, after all, created by humans and there is no reason to think that we, in this regard, are more powerful collectively than the signalman is individually. Is it because *propositions* are the real truth-bearers, so that the *real* language is not created by humans, and, being abstract and disembodied, the *real* language of propositions is divine and therefore able to do what is impossible for mere

mortals? Making that kind of move is certainly a well-tested method for forcing a desired element of perfection into a worldview where it otherwise seems lacking.

Is the difference that English is semantically closed, and that semantic closure can force states of affairs to both obtain and fail to obtain, or neither obtain nor fail to obtain? The entire communication system in scenario 2, consisting of the court sending Kuvata instructions and him trying to follow them, is also semantically closed (if English is), as the court can use the full resources of English to convey a criterion for signalling to the signalman.

Is the difference that English is compositional? If Kuvata didn't receive individual instructions for a finite number of instants of time, but instead received an instruction manual containing recursive rules according to which he could deduce what was expected of him at each instant of time during the day by applying the rules to, say, a number that represents the instant of time in some chronometric system; and it was a result of *those* rules that the ought-to-light condition for Kuvata at sundown is that Kuvata does not light a fire at sundown, would he then both light the fire and not light the fire (or neither light it nor not light it)?

There are no relevant differences between the two scenarios. The courtiers who interpret the signal sent from the signalman in accordance with its ought-to-send conditions are expecting too much from him. They have imposed obligations on him that he cannot live up to. Glut and gap theorists, who think that each sentence is true iff its ought-to-be-true conditions are satisfied, are expecting too much of language. The language community has imposed obligations on the language that it cannot live up to.

2 Initial clarifications and overview

I hope I have made my point vivid. There is still a lot of work to be done to make my point precise, avoid various ways it can be misunderstood, and argue for it in detail. Those are the tasks of the rest of this paper.

Let me first clarify what I mean by the principle that any state of affairs either obtains or fails to obtain. What I intend to exclude are "both/and" cases and "neither/nor" cases, not "in between" cases. Say that the degree to which a state of affairs obtains is represented by a real number x between 0 and 1, and that the degree to which that state of affairs fails to obtain is represented by a real number y between 0 and 1. I *happen to* think that only the two options $(x, y) = (1, 0)$ and $(x, y) = (0, 1)$ are possible when "state of affairs" is properly understood. But that claim is controversial. What I intend to deny, for the purposes of this paper, is that it is possible for $x + y$ to be different from 1. That is, I intend to deny gap theory that allows for $x + y = 0$ and glut theory that allows for $x + y = 2$, not the possibility

of vagueness in a state of affairs obtaining where, e.g., $x = y = 0.5$. For present purposes, that sort of situation is just a distracting complication. If you think that there is vagueness in the case of the state of affairs of the liar being true, replace, in this paper, the liar with the definite liar: the sentence "the definite liar is not true to degree 1". And if you think that vagueness might be relevant to the case of the state of affairs of Kuvata lighting a fire (say if he lights just a few of the available branches), then change the story so that the court has stipulated a precise criterion for when a fire counts as lit.

A fact is an obtaining state of affairs, and a state of affairs is a way the world might be. The last half of that explanation is of course very imprecise, but it will do for present purposes.

Clarification is also needed with respect to what exactly I am arguing against. Most of that work is done in section 5, using a distinction between the truth property and the truth predicate, but some initial and tentative remarks are in order. Above, I characterized gap theorists as those according to whom the liar is neither true nor not true. Actually, the gap theorists I have in mind would never allow themselves to say "λ is neither true nor not true" because that claim itself is not part of their theory. But I disagree with them, so *I* think the above sentence is a good way to characterize their position. (Compare this with a situation in which someone holds the view that it is impossible to make sense in English. *I* would not hesitate to characterize their view in the way that I just did.)

The simplest and most extreme type of gap theorist would be obligated to be a quietist on the issue of the truth value of λ, on pain of hypocrisy. For that reason, I cannot point to any actual person who holds that view. Slightly less extreme gap theorists, on the other hand, are able to convey their views. Those who have written about their position have suggested ways out that allow them to write about it: for example, that they can *deny* λ and also *deny* the negation of λ; or, they might say that they do not "accept that there is a truth of the matter" (Field 2008, 70).

Dually, the glut theorists I have in mind might very well accept what *I* intend to be a denial of their position: that λ is *not* both true and not true. It is just that they would also accept that λ *is* both true and not true. See, e.g., Priest (2006a), especially section 2.7, and Priest (2006b), especially section 4.9.

As mentioned above, I will return to this issue in section 5, as there are some delicate distinctions that need to be made first. The rest of this paper can be briefly outlined as follows. In section 1 I considered and rejected some candidates for a relevant difference between the two scenarios. A stronger argument for the claim that there are no relevant differences is given in section 3. In section 4, I will explain what I meant when I wrote that

the truth value of λ is misleading, and in what sense it "ought" to be the opposite of what it is. Section 6 contains a "prequel" to scenario 1 and an account of how language's origin as a human tool of communication is relevant to the liar. Section 7 concludes the paper with a discussion of the dialectical situation.

To foreshadow the clarifications I will make later: I will claim that while the meaning of indicative sentences is given by truth conditions, there must be exceptions to the general compositional rules that typically determine those truth conditions.

3 The Inclosure Schema

Above I have applied Priest's (1994) own Principle of Uniform Solution: similar paradoxes should be given similar solutions. It is clear that the solution in the case of scenario 2 is that Kuvata fails to represent the facts as they are because the set of sent signals must include some that misrepresent the facts (relative to the interpretation of the signals by someone who fails to realise the problem with the instructions given to the signalman), and I therefore concluded that English fails to represent the facts as they are because the set of true sentences must include some that misrepresent the facts or the set of non-true sentences must include some that misrepresent the non-facts (relative to the interpretation of English sentences by someone who fails to realise the problem with the "instructions" we give to each other in the form of intended compositional language conventions).

But are the paradoxes really similar? Is the analogy apt? I have offered some reasons for thinking so, and to my mind they are sufficient. However, someone who disagrees with my conclusions regarding the liar may think that their points of dissent in that regard themselves exactly constitute relevant points of disanalogy. Accusations of begging the question may ensue. A standoff could result.

Fortunately, there is a method for determining whether something is similar to the liar—in the sense relevant to the Principle of Uniform Solution—that is independent of views about what the correct solution to the liar is. The method is to check whether that something fits into the Inclosure Schema, also formulated by Priest (1994, 28; 1995, 147). While I disagree with Priest's dialetheism, I am convinced by his arguments that the Inclosure Schema is an adequate test. So let us apply it. The Inclosure Schema looks like this:

There are two properties φ and ψ and a (possibly partial) function δ such that

1. $\Omega = \{y \mid \varphi(y)\}$ *exists,* $\psi(\Omega)$*, and*

2. *if x is a subset of Ω such that $\psi(x)$, then*

   (a) $\delta(x) \notin x$ *and*

   (b) $\delta(x) \in \Omega$

In order to fit scenario 2 into the schema, we let y range over instants of time and x range over sets of instants of time, and we let δ map sets of instants of time to instants of time. Further, for any condition ϕ, let $\langle\phi\rangle$ be the unique instant of time t such that Kuvata has been instructed by the court to light the fire at t iff ϕ is satisfied (if there is such a t; otherwise, it is not defined). Then, we can specify the two properties and the partial function needed:

- $\varphi(y)$: y is an instant of time when a fire is lit by Kuvata
- $\delta(x) = \tau$ where $\tau = \langle\tau \notin x\rangle$ (if there is such an instant τ; otherwise, it is not defined)
- $\psi(x)$: the court has given instructions that imply that $\delta(x)$ is defined[1]

The first half of condition 1 is satisfied, because $\Omega = \{y \mid \varphi(y)\}$ is the set of instants of time when a fire is lit by Kuvata, so it exists. The second half is as well: the court has given instructions that imply that sundown is the unique instant of time t such that the court has instructed Kuvata to light a fire at t iff $t \notin \Omega$, so $\psi(\Omega)$.

To show that condition 2 is satisfied for the liar paradox, Priest makes use of the T-schema. The analogue here is a principle we can name *Perfection*: for any condition ϕ for which $\langle\phi\rangle$ is defined, $\langle\phi\rangle \in \Omega$ iff ϕ is satisfied (that is, Kuvata is perfect: he lights the fire at a given instant if and only if the ought-to-light condition associated with that instant is satisfied).

Now, assume that a is a subset of Ω such that $\psi(a)$. Then we have

$$\begin{array}{lcll} \delta(a) \in a & \Rightarrow & \langle\tau \notin a\rangle \in a & \\ & \Rightarrow & \langle\tau \notin a\rangle \in \Omega & \\ & \Rightarrow & \tau \notin a & \text{(by Perfection)} \\ & \Rightarrow & \delta(a) \notin a & \end{array}$$

Hence, $\delta(a) \notin a$, so condition 2a is satisfied. Moreover, it follows that $\tau \notin a$, and hence by Perfection that $\langle\tau \notin a\rangle \in \Omega$, i.e., $\delta(a) \in \Omega$, so 2b is also satisfied.

[1] In the case of the liar paradox, y ranges over sentences, x ranges over sets of sentences, $\varphi(y)$ means that y is true, $\psi(x)$ means that x is definable, and δ is a function defined by diagonalisation in such a way that if x is definable then $\delta(x) = \alpha$ where $\alpha = \langle\alpha \notin x\rangle$, where $\langle\phi\rangle$ now is a sentence that is true iff ϕ is satisfied.

Since $\psi(\Omega)$, a contradiction follows: $\delta(\Omega) \notin \Omega$ and $\delta(\Omega) \in \Omega$.

Of course, Perfection is not true, but as Priest (2010, 360–361) writes, "[i]t is not required that the arguments entailing the conditions be sound." The principles used in them just have to be prima facie plausible, and Perfection is plausible, *until* one is confronted with scenario 2 or something similar, just like the T-schema is plausible, *until* one is confronted with scenario 1 or something similar.

The conclusion is that scenario 2 does fit into the Inclosure Schema, implying that the Principle of Uniform Solution applies to the set of the two scenarios.[2] Since the claim about scenario 2 made in section 1 is obviously correct, we can conclude that the claim about scenario 1 is as well.

In addition to providing a formal demonstration that scenario 2 satisfies Priest's criterion for being relevantly similar to scenario 1, I would like to explain more informally why I think scenario 2 is particularly well suited to teaching us something about scenario 1. That is because there is an ideal combination of similarities and differences between the two. Among the similarities is, first, the fact that they both concern a language: the sentences in scenario 2 are the potential acts of lighting a fire at a specific time and place and the potential acts of refraining from doing so. Second, both languages contain self-referential sentences, namely λ and its negation (among others) and the potential acts of lighting a fire or not at sundown. And third, they are both systems devised and used for the purpose of communication by fallible humans. The difference is that the paradoxical sentence in scenario 1 makes reference to *truth*, which is a difficult concept to understand—causing philosophers to formulate and defend almost any imaginable position regarding it—while the paradoxical sentences in scenario 2 instead refer to a fire being lit, which could hardly be more concrete. This concreteness makes the solution to the paradox of scenario 2 obvious. As this is the only significant difference between the two scenarios, there is only a degree-of-obviousness difference between the solutions, not a substantial one: in both cases, the human-created systems are prevented from being perfect by the basic metaphysical principle that states of affairs either obtain or do not.

There is a bonus advantage to having gone through the trouble of showing that scenario 2 is an instance of the Inclosure Schema: namely, that the precise details of the analogy between the two scenarios were revealed. The ranges of the variables and the interpretations of the properties and the function in the Schema, plus the function $\langle\cdot\rangle$, provide a mapping between the scenarios. This is important because the analogy is potentially confusing. It is *not* the most natural analogy between the two scenarios, that is, it

[2] As is well known, there are also many other paradoxes that satisfy the Schema, but to keep things as simple and brief as possible, I will not generalize my conclusions regarding the liar to those other paradoxes.

does not in every case map an aspect a_1 of Kuvata's language to the aspect a_2 of English that is most similar to a_1. It being a language that contains declarative sentences, it would make perfect sense to talk about truth in connection with Kuvata's language: some potential acts of lighting a fire and refraining from doing so at specific instants of time are true, and others are not. Hence, the most natural analogy between the two languages would map the property of being a true potential act of lighting a fire or a true potential act of refraining from lighting a fire to the property of being a true sentence (and Kuvata lighting a fire to someone uttering an English sentence). However, nothing interesting about the liar could have been learned from considering that analogy. Instead, I have mapped Kuvata lighting a fire to truth, which allows us to transfer insights from an uncontroversial to a controversial domain. Fortunately, there is no requirement that, for the Principle of Uniform Solution to apply to a pair of two paradoxes, they must fit into the Inclosure Schema in accordance with a natural analogy. For instance, Priest (1994) fits Berry's paradox in by letting $\varphi(y)$ mean "y is a natural number that can be defined in less than 99 words" even though Berry's paradox is also about English, implying that truth is available for an analogy to map to.

4 Misleading language

Given the claim about scenario 1, the truth conditions of λ cannot be what they seem to be, because then λ would be true iff λ is not true. I do not have a thesis about what the actual truth conditions of the liar are. I am only doing two things in this paper. First, I am ruling out two extreme positions concerning the liar. Second, I am making a point about language being potentially misleading that should lower our expectations regarding any thesis about what the truth conditions and truth value of the liar are. Many proposed solutions to the paradox try to convince us of two things: namely, that truth values are distributed over the sentences of a language that contains the liar in some specific way, and that such distribution is entirely natural and satisfying. My point is that we have to acknowledge that *no* possible distribution of truth values is entirely natural and satisfying, and so neither is the actual distribution, whatever it happens to be. We have demanded so much of language that something was bound to go wrong.

My claim that truth misrepresents the facts as they are may seem as baffling as the liar paradox itself. How could *truth* possibly depart from correct representation of the facts?! Well, it only does relative to a certain interpretation of the true sentences; and the actual truth conditions of some sentences must be different from what we think they are, when we interpret the true sentences that way. Kuvata only misrepresents the facts relative to a certain way of interpreting his signals. The actual conditions under which

he lights a fire are different from what the courtiers think they are, when they interpret his signals in such a way that the facts are misrepresented to them. Just to mention one of many possibilities, Kuvata might default to not lighting the fire when he receives the kind of paradoxical instructions exemplified in scenario 2. That is, when the court gives an order of the form *Kuvata lights the fire at t iff the fire is not lit at Mount m at t*, the actual condition might be that *Kuvata lights the fire at t iff the fire is not lit at Mount m at t and Mount m is different from Mount Suomalainen.* Similarly, actual truth conditions of paradoxical sentences must come apart, in some way, from what sentences carry on their sleeves. So a detailed theory about the distribution of truth values should not be admonished for implying this necessity.

Hence, my claim that truth misrepresents the facts is not this claim: "Truth misrepresents the facts relative to the actual truth conditions of the true sentences." *That* would not only be baffling; it would be impossible.

We can thus distinguish between two versions of the T-schema, listed here together with the corresponding schemata concerning Kuvata lighting a fire:

(T1) ϕ is true iff the actual truth conditions of ϕ are satisfied

(T2) ϕ is true iff the truth conditions of ϕ according to the default compositional rules are satisfied

(F1) Kuvata lights a fire at t iff the actual conditions for Kuvata lighting a fire at t are satisfied

(F2) Kuvata lights a fire at t iff the conditions for Kuvata lighting a fire at t stipulated by the court are satisfied

(T1) and (F1) must hold by necessity but do not lead to paradox. (T2) and (F2) may fail, and when that happens it leads to language being misleading. (T2) fails for λ because, according to our default compositional rule for negation, a sentence and its negation have opposite truth values, and according to our default compositional rule for the truth predicate, a sentence and another sentence consisting of the truth predicate applied to a name of the former have the same truth value.

The generalization of (F2) is what I called "Perfection" in the previous section, and (T2) is the version of the T-schema that Priest relies on.

The right-hand sides of (T2) and (F2) are the precise versions of what I expressed above, in a suggestive but imprecise manner, using the words "ought" and "obligations". What I called Kuvata's "ought-to-light" conditions are merely prima facie obligations, not all-things-considered obligations. And the expression "ought-to-be-true" was merely metaphorical.

5 The truth property and the truth predicate

The opponents I have argued against are the radical gap theorist who thinks that one and the same sentence can both fail to be true and fail to be not true and the radical glut theorist who thinks that one and the same sentence can be both true and not true. I have no beef with anyone who simply claims that some sentence is not true while that sentence prefaced with, e.g., "It is not the case that" is also not true, or, dually, claims that both sentences are true. The kind of argument used in this paper cannot be employed to show that someone of that persuasion is wrong, as the following scenario demonstrates.

Scenario 3: One year, Kuvata receives the following instructions: "Each day in the coming year, if the signalman on Mount Suomalainen does not light a fire at sundown—and only if he does not—you must light a fire at sundown. Also, if the signalman on Mount Suomalainen does light a fire at sundown—and only if he does—you must light a fire one sandglass after sundown."

Just as in scenario 2, Kuvata either lights the fire at sundown or he doesn't, and whether it's one or the other, that signal will be misleading. The point of this new scenario is that one sandglass after sundown, Kuvata can light a fire or refrain from doing so, and he is not forced to do the opposite of what he did at sundown even though the conditions stipulated by the court for the two instants are opposite. (In fact, if Kuvata wants to minimise the number of misleading signals he sends, after he has been forced to abandon the ideal of never sending misleading signals, he will light the fire one sandglass after sundown if he lit the fire at sundown, and not if not.)

This has consequences. If we combine what I have argued for (prior to this section) with what I have made explicit that I will not rule out (in this section), it follows that λ cannot be both true and not true, but that the glut theorist who claims as much by asserting both "λ is true" and "λ is not true" might be using only true sentences to do so (and vice versa for the gap theorist)![3]

However, that just serves to underscore my point: true sentences can be misleadingly true (or non-true sentences can be misleadingly non-true). The way to avoid being misled into thinking that I have just disavowed one of

[3]Thus, my position has similarities to (but is not identical with) Mares's (2004), who rejects *metaphysical dialetheism* while accepting the possibility of *semantic dialetheism*; and to Tahko's (2009), who tries to formulate and defend a metaphysical version of the law of non-contradiction. Further, I *think* that the positions I am distinguishing between are represented by actual philosophers as follows. Radical glut: Priest (2006b), radical gap: Field (2008), moderate glut (as just explained): Beall (2009), moderate gap: Kripke (1975). However, differences of vocabulary and differences of opinion about related matters make this question of interpretation difficult.

my main theses is to distinguish between the truth property and the truth predicate. There is a truth property: it is the (complex) property of being a sentence to which truth conditions have been assigned (by convention) and those truth conditions being satisfied. For a sentence ϕ to be true is for ϕ itself to have the truth property, not for the sentence "ϕ is true", containing the truth predicate, to have the truth property (unless ϕ is the truth teller). The state of affairs of ϕ being true can not both obtain and fail to obtain, for it is just *one* state of affairs, like Kuvata lighting a fire at one particular instant of time is. In contrast, the string of words "ϕ is true" having the truth property and the string of words "ϕ is not true" having the truth property are two different states of affairs and they might very well obtain or fail to obtain independently of each other (even though both obtaining or both failing to obtain would imply that one of them is misleading), just as Kuvata can make independent decisions about lighting fires at two different instants of time.[4]

Let me elaborate a bit. Both Kuvata and English attempt to represent whether certain states of affairs obtain or not by letting other states of affairs obtain or not. We can distinguish between states of affairs that represent and states of affairs that are represented. Typically, they are completely different states of affairs: the state of affairs of the sentence "Mark is happy" having the truth property represents, and the state of affairs of Mark being happy is represented. In the case of self-reference, the same states of affairs may play both roles. However, even in those cases, we can distinguish between the two *roles* of the same state of affairs, and that is the distinction I am appealing to here. The state of affairs of the sentence "λ is not true" having the truth property is supposed to represent; and it is the state of affairs of the sentence "λ is not true" having the truth property that it is supposed to represent. The state of affairs of the sentence "λ is true" having the truth property is also supposed to represent, and it is also the state of affairs of the sentence "λ is not true" having the truth property that it is supposed to represent (just in a different way). It is the fact that there are two different states of affairs that represent that allows this kind of "weak glut" or "weak gap". There is no glut or gap in states of affairs, just a violation of the intention behind the negation as a tool for representation: namely, that exactly one of these sentences "ought" to have the truth property.

So, to be clear, I am not arguing against *any* specific thesis about which

[4]While there is much I can agree with in Chihara (1979), the distinction between the truth property and the truth predicate can be used to show what is missing from his account. Chihara compares truth to *glub*, defined by "an animal is a glub if, and only if, it is not a mouse; and it is not a glub if, and only if, it is neither a mouse nor different from itself" (592), and argues that they are similar: in particular, that they are both "inconsistent concepts". But there is also a very important difference: the truth predicate is intended to correspond to what is a perfectly fine property (and only prevented from always doing so by self-reference). In contrast, there is no glub property.

sentences have the truth property, which sentences do not, or which combinations of having and lacking the truth property are possible for any set of sentences. In particular, I am not ruling out any thesis about which sentences containing the truth predicate have the truth property, such as, e.g., "λ is true or $\neg\lambda$ is true". More generally, I am not ruling out any thesis about sentences of particular *forms* having and lacking the truth property—say, for instance, the form "ϕ and not ϕ". That is, I am not arguing for or against classical logic.[5] And since (F2) can fail in either direction, depending on what Kuvata decides to do, I am not taking a stand about the direction of failure in the instance of (T2) where ϕ is λ, i.e., whether "capture" or "release" fails for this version of the T-schema. I am just claiming that every sentence either has or fails to have the truth property, and not both. That is merely a consequence of the metaphysical principle mentioned above, but nevertheless sufficient to imply (given the existence of the liar or certain similar sentences, together with our naive, compositional expectations about their truth conditions) something surprising: that some sentences have or lack the truth property in a way that is misleading.

I can only hope I am not *using* too many such sentences in this attempt to communicate about a communication problem—and, to the extent that I am, that the reader is not as easily misled as the courtiers who interpret all received signals strictly in accordance with the instructions given to the signalmen.

6 Scenario 1, extended version

In Scenario 2, at the beginning of this paper, I described in some detail how the naive courtiers came to interpret Kuvata's lighting of a fire at sundown as indicating that Kuvata did not light a fire at sundown. In contrast, my description of Scenario 1 was extremely brief. I will now make up for that by describing how I think we ended up naively interpreting λ in the way that we tend to.

Prequel to scenario 1: Until 100,000 years ago,[6] no sentence was true. There were lots and lots of objective and mind-independent facts, of course—but no sentence was true. That changed when some tribe of Stone Age people realized that they needed a tool with which to communicate about matters of fact, and that they could create that tool by assigning truth conditions to strings of sounds through conventions. There are different ways that a

[5] However, you might call what I am defending a classical theory of states of affairs. And it does imply this weak thesis concerning classical logic: If a given set of declarative sentences contains only sentences with non-misleading truth conditions, then classical logic holds for that set.

[6] It should be equally clear that there are certain details in the following that I cannot be certain of, and that their accuracy does not matter.

language society may attach truth conditions to such sentences by convention, and initially this tribe did it in the most primitive way: they instituted individual conventions for individual sentences.[7]

Despite this primitivity, this gave the tribe a useful tool. If one member of the tribe noticed that some state of affairs obtained, and one of the sentences of the tribe's language had as its truth condition that this particular state of affairs obtains, then this individual could utter the sentence, and the other members of the tribe would understand that the utterance indicated that the state of affairs in question obtained. Information would be transferred.

So, as of 100,000 years ago, some sentences were true. That is, some sentences had the truth property, because truth conditions had been assigned to the sentences by a language community, and those conditions were satisfied. However, there still wasn't any truth predicate. In general, there were no predicates, because the sentences did not have any meaningful proper parts at all.[8]

That made the language inefficient: there were only finitely many sentences—one for each convention the tribe managed to institute—and they had to be learned one by one. So 99,000 years ago, the tribe came up with a better way to attach truth conditions to sentences by convention: indirectly, through the use of compositional rules. Instead of having conventions for individual sentences, the tribe now created conventions for how words contributed to the truth conditions of different sentences. That made for a much more efficient language containing a potential infinity of sentences, which allowed the members of the tribe to communicate about matters that they had not anticipated the need to be able to communicate about in advance. It was also easy to learn, and easy to use.

Such a compositional language is *particularly* easy to learn and to use if the compositional rules are completely uniform and admit of no exceptions. But of course, this pragmatic uniformity desideratum does not imply that there *cannot* be exceptions. The English language community could decide, collectively, that from this day forward the sentence "grass is green" should have as its truth condition that snow is pink; and we could make that change without changing the truth conditions of *any* other sentence, including, e.g., "grass is not green". It would be a silly thing to do, as there would be a cost to simplicity without any benefit; but we *could* do it.

[7] See Lewis (1969) for an account of how conventions can be instituted in the absence of antecedent language conventions, as well as a more in-depth account of what it is for a sentence to have the truth property.

[8] To make it explicit: I am here contradicting deflationists like Beall (2009), who claim that there is no truth property independent of the truth predicate. I am also contradicting those who claim that truth is primitive (e.g. McGinn (2000)); on the contrary, it is a complex property that depends both on facts about social conventions and facts about whatever those conventions have made the given sentence be about.

If we did, then neither "grass is green" nor "grass is not green" would have the truth property: this pair would, in a superficial sense, constitute a gap. It is in this same superficial sense that I allowed for the possibility of a gap for the pair "λ is true" and "λ is not true" in the previous section.

When the first indicative sentences were introduced 100,000 years ago, there was a new kind of fact in the world, namely semantic facts about certain sentences having the truth property. Then, 98,000 years, ago the tribe decided that they would also like to be able to communicate about those facts. Therefore, the tribe introduced the truth predicate. The idea was that a sentence consisting of the truth predicate applied to the name of a sentence should have the truth property iff the named sentence had the truth property. This is a useful compositional rule—especially if it is uniform.

Slightly earlier the tribe also introduced a negation operator, and in this case the idea was that the negation of a sentence should have the truth property iff that sentence fails to have the truth property. Again a very useful compositional rule—especially if it is uniform.

At the moment the tribe also created the necessary linguistic resources to allow the truth predicate to be applied to a name of the negation of the resulting sentence, we arrived at the beginning of the original description of scenario 1. There now was, and still is, a liar sentence: i.e., a sentence such that *if* the compositional rules were completely uniform, *then* the sentence would have the truth property iff it does not. So a language user who naively assumes that the compositional rules must be completely uniform would interpret the sentence as having that impossible truth condition.[9] But that must be wrong. There is another sentence that both (1) consists of the truth predicate applied to a name of the liar sentence and (2) is the sentence negated by the liar sentence. Hence, there must be an exception to either the compositional rule for the truth predicate or the compositional rule for negation. If an exception to a compositional rule *can* happen simply because of a decision by the language community, as in the "grass is green" example above, then it *must* happen when the alternative is a state of affairs that both obtains and fails to obtain, which there would be here: the state of affairs of the liar sentence having the truth property would both obtain and fail to obtain.[10]

How does this story end? What are the actual truth conditions of the liar

[9]Such a language user would also misinterpret "he kicked the bucket" and most other idioms.

[10]Such failures are consistent with English being computationally compositional; a property that may be used to explain the possibility of language acquisition. Computational compositionality does not require that the truth value of $\neg\phi$ or $T(\ulcorner\phi\urcorner)$ be a function of the truth value of ϕ, only that the meaning of $\neg\phi$ and $T(\ulcorner\phi\urcorner)$—together with the meaning of any other of the potential infinity of sentences—can be "worked out" from knowledge of finitely many rules.

sentence? Well, one option is that it does not have any—that we have not managed to institute a convention for this sentence.[11] However, while that is consistent with the rest of the story, it does not follow from it. Another option is that, say, a widespread conservative psychological tendency among the members of the language community makes us default to non-truth when the usual compositional rules fail, resulting in the actual truth conditions of the liar sentence being that the liar is not true *and* that it is grounded. A third option is that a widespread psychological tendency to insist on a right to express ourselves about anything has resulted in a language convention that allows two sentences that seem to contradict each other to both have the truth property. There are also many other options: a great deal of the proposals that have been put forward in the literature for how to solve the paradox can be reinterpreted as descriptions of possible but non-necessary language conventions. Finding out which one is the actual language convention requires empirical investigation.[12]

It is true of all language conventions that knowledge about them is a posteriori. Some parts of that knowledge are easy to come by, and other parts are difficult. At the easy end of that spectrum there is knowledge possessed by most three-year-old members of the language community, and at the difficult end, insights that still elude the most gifted linguists. Our questions about the convention that governs the truth conditions of the liar sentence—whether there is one, and if so, what it is—are clearly at the difficult end. It would be extraordinarily difficult for an empirical linguist to investigate them. A survey would most likely result in data for which it would be close to impossible to separate the signal from the noise—the signal in this case being evidence of actual language conventions, and the noise being the re-

[11] If so, one might question whether it should be called a "sentence". However, I am using that term in a wide sense that tracks whether there is a compositional *intention*. I did something similar with "truth predicate", "negation (operator)", and "liar sentence" above.

[12] A comparison with Russell's barber may be helpful here. Usually, the solution to that paradox is expressed simply as "there is no such barber". While this also fits with our other cases—there is no sentence that has as its actual truth conditions that it is not true; there is no instant of time t such that Kuvata actually lights a fire at t iff not—it is more illuminating if we think of the paradox of the barber slightly differently. Instead of defining the barber in the usual way, define him to be someone living in a village where all the inhabitants *share a sincere intention* to follow a convention according to which the barber shaves all and only those who do not shave themselves. Such a barber could exist. Then the solution is, instead, that there is a difference between the intended convention and the actual convention, which might be, e.g., that the barber shaves all and only those who do not shave themselves, except that he also shaves himself. All, or most of, the villagers may be ignorant of the fact that there is (and must be) a discrepancy between the intended and the actual convention, and that the latter is not uniquely determined by the former, but requires further psychological determiners. All of this is closely analogous to the situation of the liar. (It is *slightly* less closely analogous to the situation of Kuvata, but only because I allowed commands to replace conventions in that scenario in order to make it more concrete.)

spondents' folk-philosophical theories about self-reference. Nevertheless, it *is* an empirical question: the answer depends on contingent facts about our society.

7 Conclusion

The court tasked Kuvata with representing certain facts by the lighting of fires, and attempted to do so by assigning light-a-fire conditions to instants of time through decrees. But when they attempted to effect the representation of certain facts about the lighting of fires by having specific instants be ones where fires were lit, the light-a-fire conditions that they attempted to assign could not become the actual conditions under which Kuvata lit a fire.

Human societies have tasked themselves with representing facts by making sentences true, and attempted to do so by assigning truth conditions to sentences through exception-free compositional conventions. But when they attempted to effect the representation of certain facts about truth by having specific sentences be true, the truth conditions that they attempted to assign could not become the actual truth conditions.

There are certain differences between the two systems of representation. First, truth is a more complex property than that of being ablaze. Second, after the court has issued its instructions, Kuvata has to take physical action to create the representation; but once a language society has instituted a convention for the truth conditions of a sentence, no further act is required before that sentence becomes true (or not). Third, while it is easy to find out whether Kuvata has lit a fire (just look out of the windows of the palace!), it is in many cases difficult to ascertain whether a given sentence is true: epistemic access both to the facts that determine whether the truth conditions are satisfied, and to the facts that determine what the truth conditions *are*, can be impaired. But those differences should not mislead us into thinking that the second system of representation can achieve a level of perfection that violates the metaphysical principle that any state of affairs either obtains or not, and not both.

I will close by mentioning two ways that someone might disagree with this conclusion due to background assumptions that I have not argued against. First, it should be clear from the above that I take language to be a cultural artifact, and as such, created in time and subject to human limitations. Someone who believes in abstract and mind-independent propositions may feel my account misses the target, because such propositions are assumed to be the primary bearers of truth. To my mind, that would be to multiply entities beyond necessity; but a discussion about that must be declared to be beyond the scope of this paper. So officially, the claims I have made above

are conditional rather than categorical, with linguistic nominalism being the condition.

Second, I have assumed that there are no good reasons different from the liar (and the other semantic paradoxes) for rejecting the metaphysical principle that any state of affairs either obtains or fails to obtain. But Priest (2006b, Chapters 11–13), for instance, believes that transitions in the world involve gluts, and that a parliament passing an inconsistent law can make something both legal and not legal. For someone who *already* believes that gluts are possible, and that human beings have the power to create them, it seems quite reasonable to conclude that our inconsistent compositional intentions did just that. But in the absence of such independent justification, my claim is that the liar sentence does not provide much of a reason to change one's mind, for there is a satisfactory alternative account that doesn't require a revision of beliefs about basic metaphysical principles: we have collectively failed in the same way that Kuvata and the courtiers failed.

Acknowledgments

I gratefully acknowledge the help and feedback I have received from Thomas Brouwer, Andreas Fjellstad, Martin Rønberg, Naja Rønberg, Øystein Linnebo, Filippo Ferrari, Gerard Hough, Crispin Wright, Robert Trueman, Shay Logan, James Shaw, Eirik Gjerstad, Eric Epstein, Luisa Venancio, Silvia Jonas, Olla Solomyak, Georgie Statham, Sharon Berry, Elena Fiecconi, Elvira Di Bona, Claire Benn, Daniel Telech, Leora Katz, and the anonymous reviewers, as well as audiences in Cambridge, Copenhagen, Aarhus, Ascona, Warwick, Krakow, Bologna, Jerusalem, Witwatersrand, Salzburg, and Vichy. The title of this paper was inspired by those of Eldridge-Smith's (2011) paper about the Pinocchio paradox and Beall's (2011) reply.

References

Beall, J. (2009). *Spandrels of Truth*. Oxford University Press.

Beall, J. (2011). Dialetheists against Pinocchio. *Analysis 71*, 689–691.

Chihara, C. (1979). The semantic paradoxes. *Philosophical Review 88*, 590–618.

Eldridge-Smith, P. (2011). Pinocchio against the dialetheists. *Analysis 71*, 306–308.

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

Kripke, S. (1975). Outline of a theory of truth. *The Journal of Philosophy 72*, 690–716.

Lewis, D. (1969). *Conventions*. Harvard University Press.

Mares, E. (2004). Semantic dialetheism. In G. Priest, J. Beall, and B. Armour-Garb (Eds.), *The Law of Non-Contradiction*. Oxford University Press.

McGinn, C. (2000). *Logical Properties*. Oxford University Press.

Priest, G. (1994). The structure of the paradoxes of self-reference. *Mind 103*, 25–34.

Priest, G. (1995). *Beyond the Limits of Thought*. Oxford University Press.

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

Priest, G. (2006b). *In Contradiction* (second ed.). Oxford University Press.

Priest, G. (2010). Badici on inclosures and the liar paradox. *Australasian Journal of Philosophy 88*, 359–366.

Tahko, T. (2009). The law of non-contradiction as a metaphysical principle. *Australasian Journal of Logic 7*, 32–47.