Reflective Knowledge and the Nature of Truth

José L. Zalabardo
University College London

BIBLID [0873-626X (2016) 43; pp. 147-171]

Abstract
I consider the problem of reflective knowledge faced by views that treat sensitivity as a sufficient condition for knowledge, or as a major ingredient of the concept, as in the analysis I advance in *Scepticism and Reliable Belief*. I present the problem as concerning the correct analysis of SATs — beliefs to the effect that one of my current beliefs is true. I suggest that a plausible analysis of SATs should treat them as neither true nor false when they ascribe truth to a non-existent belief. I argue that the problem is inescapable if we construe SATs as ascribing the property of truth to a belief. Deflationism manages to avoid the problem of reflective knowledge, but it does so by violating alethic priority — the principle that our account of representation must be built on our account of truth. I argue that we can avoid the problem of reflective knowledge while preserving alethic priority with a pragmatist account of truth — according to which truth is explicated in terms of the rules that govern the practice of assessing judgments and related items as true or false.

Keywords
Reflective knowledge, truth, pragmatism, meaning, sensitivity.

1 Sensitivity and reflective knowledge

If a subject *S* believes a proposition *p*, let's say that *S*'s belief that *p* is *sensitive*, or that *S sensitively believes* that *p*, just in case the following subjunctive is true: if *p* were false, *S* wouldn't believe that *p*.

Let me refer as *self-ascriptions of truth* (SATs) to beliefs of the form *my belief that p is true*, where *p* is a proposition that I believe now.

On the face of it, SATs can't be sensitive. If I believe that *p*, then I will believe that my belief that *p* is true whether or not it is as a

Disputatio, Vol. VIII, No. 43, November 2016
Received: 19/07/2016

matter of fact true. So long as I believe that *p*, my propensity to believe that my belief that *p* is true will not be affected by the truth value of *p*.

Consider now the view that sensitive belief is a necessary condition for knowledge, to which I'm going to refer as the *Sensitivity Constraint* (SC):

> SC: If *S* doesn't believe that *p* sensitively, then *S* doesn't know that *p*.

If SATs can't be sensitive, then, according to SC, they can't be knowledge either. I can never know that my beliefs are true. This is an implausibly radical sceptical outcome. Since SC appears to make this outcome unavoidable, we have a very powerful reason for rejecting SC. This is what's come to be known as *the problem of reflective knowledge* for SC.

The problem has been developed in some detail by Jonathan Vogel. He uses the following example:

> You see your long-time friend Omar, who is a perfectly decent and straightforward sort of person. Noticing his shiny white footwear, you say, "Nice shoes, Omar, are they new?" Omar replies, "Yes, I bought them yesterday." I think the following things are true:
>
> (10) You know Omar has a new pair of shoes.
>
> (11) You know that your belief that Omar has a new pair of shoes is true, or at least not false. (Vogel 2000: 609-10)[1]

Vogel argues that SC is incompatible with (11), since you don't sensitively believe that your belief that Omar has a new pair of shoes is true, or at least not false:

> As things actually are, you believe that your belief that Omar has new shoes is not false. What if it were? If somehow your belief that Omar has a new pair of shoes were false, you would still believe that your belief was true, not false. The alternative is hard to fathom. It is difficult to conceive of your not believing that something you believe is true, whenever the matter happens to cross your mind. So, if your belief that

[1] Vogel had already raised the problem in Vogel 1987: 203. See also DeRose 1995: 22-23, Sosa 1999: 145.

> Omar has new shoes were false, you would still believe that your belief was true, not false. (Vogel 2000: 610)

He spells out the argument using a particular analysis of SATs. If O stands for the proposition that Omar has a new pair of shoes, and $B(p)$ for the proposition that you believe that p, then, on Vogel's analysis, the proposition 'your belief that Omar has a new pair of shoes is true, or at least not false' has the following structure:

$$\sim(B(O) \wedge \sim O)$$

I.e. it is the proposition that it's not the case that you believe O but O is false, or that you don't believe O falsely.

Assume that you believe O and $\sim(B(O) \wedge \sim O)$. In order for your belief in $\sim(B(O) \wedge \sim O)$ to be sensitive, it's got to be the case that in the nearest world W in which $\sim(B(O) \wedge \sim O)$ is false you don't believe $\sim(B(O) \wedge \sim O)$. But this condition, Vogel argues, cannot be satisfied. In W, $\sim(B(O) \wedge \sim O)$ is false, and hence $B(O) \wedge \sim O$ is true. A fortiori, in W, $B(O)$ is true. But "[i]f you believe O, you believe that you do not falsely believe O" (Vogel 2000: 611). Hence, in W, you believe $\sim(B(O) \wedge \sim O)$. Therefore your belief that $\sim(B(O) \wedge \sim O)$ is not sensitive. It follows that, according to SC, you can't know $\sim(B(O) \wedge \sim O)$. Vogel finds this outcome unacceptable and invokes it to justify his rejection of SC.

Joe Salerno has recently attacked Vogel's argument for its reliance on the principle that if you believe that p, then you believe that you don't falsely believe that p. Call this principle *Reflection*. Salerno writes:

> [...] it is not obvious that believing p entails the higher-order belief that one is not mistaken in believing p. That implies that small children and other unreflective thinkers have beliefs about their own beliefs. More to the point, no contradiction flows from the assumption that there is a thinker who, for whatever reason, is able to form only first-order beliefs (i.e., beliefs that do not have the concept of belief as part of their content). (Salerno 2010: 75)

Vogel does indeed invoke Reflection at a crucial step in his argument, and Salerno's concerns are compelling. However, if we concede to Salerno that believing that p doesn't entail believing that you don't falsely believe that p, the argument still goes through. Bear in mind that we are investigating the epistemic status of your belief $\sim(B(O) \wedge \sim O)$. The question that we are asking is: is this belief sensitive? For this we need to consider whether you would have the belief in the nearest world W in which it is false, i.e. in the $B(O) \wedge \sim O$-world that most resembles the actual world. Vogel uses Reflection in support of his claim that in W you believe $\sim(B(O) \wedge \sim O)$. He derives this conclusion from the premise that W is a $B(O)$-world using Reflection. However, I want to argue that this is an unnecessary detour. We are assuming that in the actual world you believe $\sim(B(O) \wedge \sim O)$. It follows that you will also believe $\sim(B(O) \wedge \sim O)$ in W unless the changes that need to be made to the actual world to turn it into a $B(O) \wedge \sim O$-world would remove your belief that $\sim(B(O) \wedge \sim O)$. But there is no reason to think this. Hence we can obtain the conclusion that you believe $\sim(B(O) \wedge \sim O)$ in W, and hence that your actual belief that $\sim(B(O) \wedge \sim O)$ is not sensitive, without invoking Reflection.[2]

Kelly Becker has shown that Vogel's result depends on his specific construal of SATs — as beliefs of the form that you don't falsely believe that p, rather than of the form that you believe that p truly (Becker 2006). Suppose that we analyse the proposition 'your belief that Omar has a new pair of shoes is true, or at least not false' as having the following structure:

$B(O) \wedge O$

Assume that you believe that O and that $B(O) \wedge O$. Is the latter belief sensitive? To answer this question we need to consider whether you would have the belief in the nearest world W in which it is false, i.e. in the nearest $\sim(B(O) \wedge O)$-world.

In W, we have that either $\sim B(O)$ or $\sim O$. Let's consider each case in turn. Suppose first that in W $\sim B(O)$ — you don't believe that Omar has new shoes. Since you don't believe O in W, it follows that you don't believe $B(O) \wedge O$ either. Suppose now that in W $\sim O$ — Omar doesn't have new shoes. Now, since W is a $\sim O$-world and the nearest $\sim(B(O) \wedge O)$-world, it follows that it is also the nearest $\sim O$-world. Hence, *if your belief that O is sensitive*, in W you don't believe that O,

[2] Salerno also accuses Vogel of illegitimately invoking Closure in support Reflection (Salerno 2010: 75). Salerno is right that Closure lends no support to Reflection, but I can't see that Vogel is trying to use Closure in this way.

and a fortiori you don't believe that $B(O) \wedge O$ either.

This argument doesn't quite show that your belief that $B(O) \wedge O$ is sensitive, but the weaker result it establishes is all we need: if your belief that O is sensitive, then your belief that $B(O) \wedge O$ is also sensitive. Hence SC won't prevent your belief that $B(O) \wedge O$ from attaining the status of knowledge unless it also has the same effect on your belief that O. As far as SC is concerned, you can know that $B(O) \wedge O$ so long as you know that O.

2 Hetereogeneity

In the preceding section we have seen that the issue of the sensitivity of SATs is highly dependent on how we analyse their content. If, on the one hand, we analyse them as of the form that you don't falsely believe p ($\sim(B(p) \wedge \sim p)$), then they are necessarily insensitive.[3] If, on the other hand, we analyse them as of the form that you believe that p truly ($B(p) \wedge p$), then they will be sensitive so long as your first-order belief in p is sensitive.

It might seem, then, that advocates of SC could try to deal with the problem of reflective knowledge by taking sides with Becker and against Vogel on the question of the correct analysis of SATs. However, as Guido Melchior has pointed out, the availability of this alternative analysis simply transforms the difficulties faced by SC with respect to reflective knowledge. The new problem is this: the propositions that you believe that p truly and that you don't believe that p falsely are intuitively so similar in content that it is hard to accept that the epistemic status of your belief in one will be radically different from the epistemic status of your belief in the other. And yet, if SC is accepted, this is precisely the situation that we face. Your belief that you don't believe that p falsely cannot be knowledge, whereas, so long as your belief that p is sensitive, SC poses no obstacle to your belief that you believe that p truly also counting as knowledge. Melchior has referred to this as the *Heterogeneity Problem* (Melchior

[3] Becker has suggested that $\sim(B(p) \wedge \sim p)$ is not even the best analysis of the proposition that your belief that p is not false. He offers $B(p) \wedge \sim\sim p$ as an alternative (Becker 2006: 82). See also Salerno 2010: 77-79.

2015).[4]

I want to suggest, however, that heterogeneity doesn't sustain a cogent argument against SC. To be sure, we should expect your belief that you believe O truly and your belief that you don't believe O falsely to have the same epistemic status, and SC doesn't deliver on this expectation — *on the analyses of these propositions that we have considered*. If these analyses were correct, heterogeneity would put pressure on SC, but I'm going to argue that both analyses are incorrect.

Let's take a closer look at the relationship between $B(p) \wedge p$ and $\sim(B(p) \wedge \sim p)$. Their truth values come apart if you don't believe that p. Then $B(p) \wedge p$ is false and $\sim(B(p) \wedge \sim p)$ is true. But so long as you believe that p, their truth values are guaranteed to coincide. Then we have that $B(p) \wedge p$ is true if and only if $\sim(B(p) \wedge \sim p)$ is true if and only if p is true.

This suggests that the spurious difference in meaning that results from these analyses of SATs is produced by their diverging behaviours when the first-order belief doesn't exist — when you don't believe that p. So the question we need to ask is: what should happen to a SAT regarding your belief that p when the belief doesn't exist? According to Vogel's analysis, the SAT should be true; according to Becker's analysis, it should be false. I want to argue that neither is right. If the belief doesn't exist, the corresponding SAT shouldn't have a truth value. A SAT neither asserts nor denies the existence of the belief to which it ascribes truth. Rather, it *presupposes* its existence. Your belief that your belief that p is true or not false should be true if you believe that p and p is true, false if you believe that p and p is false, and lack truth value if you don't believe that p. A correct analysis of SATs should attribute to them this behaviour. Neither of the proposals under consideration satisfies this constraint, and the

[4] Melchior adds that the situation generated by SC is made more implausible by the fact that $B(p) \wedge p$ is stronger than $\sim(B(p) \wedge \sim p)$, since the former entails the latter but the latter doesn't entail the former: "We want an account of knowledge that allows one to know the weaker propositions d [$\sim(B(p) \wedge \sim p$] if we know the stronger propositions c [$B(p) \wedge p$]" (Melchior 2015: 483). I'm not sure how much weight this additional consideration should carry. Epistemology is full of cases in which knowing a weaker proposition seems harder than knowing one that's stronger. Take, for example, the proposition that I'm not an envatted brain and the proposition that I have hands.

appearance of heterogeneity is a consequence of this failure. In order to determine the epistemic status that SC ascribes to SATs, we need to concentrate on analyses that satisfy this requirement.

3 The predicative analysis

An analysis that satisfies our requirement is readily available if we follow the surface grammar of SATs. A SAT, like any other ascription of truth to a belief, appears to assert that the belief in question instantiates a property or satisfies a condition — the property or condition that the predicate '…is true' denotes. In a SAT, the object of predication — the belief to which this predicate is ascribed — is singled out as the referent of a definite description — as the object that satisfies the propositional function *x is a belief of mine with the content that p* ($B_p x$). Hence, if $\iota x\, Cx$ denotes *the (unique) x that satisfies propositional function C*, and T stands for the truth predicate, SATs should be analysed as of the form $T\, \iota x\, B_p x$. Now, in order for the analysis to secure the required behaviour for SATs, the definite description should be construed along Strawsonian lines, with sentences of the form $P\, \iota x\, Cx$ lacking a truth value if there isn't a (unique) object satisfying C (Strawson 1950). On this analysis, if you don't believe that p, $T\, \iota x\, B_p x$, won't have a truth value, as desired. Let me refer to this as the *predicative analysis* of SATs.

Let's consider now how the predicative analysis bears on the question of the sensitivity of SATs. As we know, in order for your belief that $T\, \iota x\, B_p x$ to be sensitive, it's got to be the case that in the nearest world W in which $T\, \iota x\, B_p x$ is false you don't believe $T\, \iota x\, B_p x$. Can your belief satisfy this condition? Notice, crucially, that the world we need to be looking at is not a world in which you don't believe that p. In such a world $T\, \iota x\, B_p x$ is not false — it lacks a truth value. The nearest world W in which $T\, \iota x\, B_p x$ is false is a world in which you believe that p but p is false. Since W is the world that most resembles the actual world in which you believe that p but p is false, and you believe $T\, \iota x\, B_p x$ in the actual world, we have to conclude that you also believe $T\, \iota x\, B_p x$ in W, since making p false without removing your belief that p does not require removing your belief that $T\, \iota x\, B_p x$. Hence in the nearest world in which $T\, \iota x\, B_p x$ is false you believe $T\, \iota x\, B_p x$. Therefore your belief that $T\, \iota x\, B_p x$ is insensitive. And in general SATs, on the predicative analysis, cannot be sensitive. It follows that, according to SC, SATs can't have the status of knowledge: I can't know that my beliefs are true. If we adopt the predicative analysis of SATs, SC renders reflective knowledge impossible.

4 Probabilistic sensitivity

The problem is not restricted to accounts of knowledge that are committed to SC. In *Scepticism and Reliable Belief*, I have defended an analysis of knowledge in which sensitivity is not a necessary condition for knowledge, but still plays a major role in the notion. I propose that non-standing beliefs (beliefs that we don't form as a result of an innate predisposition that is largely independent of input (Zalabardo 2012: 137)) can achieve the status of knowledge either by tracking the truth or through the possession by the subject of adequate evidence in their support. Following Sherrilyn Roush (Roush 2005), I construe truth tracking using, not subjunctive conditionals, but conditional probabilities. In this context, the degree of sensitivity of your belief that p is given by the probability of your belief that p conditional on p being false — $\Pr(B(p)|{\sim}p)$: the sensitivity of your belief increases as this value decreases. On my analysis, a necessary condition for your belief that p to track the truth is that you are significantly more likely to have it if it is true than if it is false. More precisely, truth tracking requires that the value of the ratio $\Pr(B(p)|p)/\Pr(B(p)|{\sim}p)$ (the *tracking ratio* of your belief) exceeds a certain threshold (Zalabardo 2012: 113-14). The probabilistic rendition of sensitivity figures in this account of truth tracking in the denominator of the tracking ratio. For any given value for the numerator, a high value for the ratio will be secured with a sufficiently low value for the denominator — i.e. by a sufficiently low probability that you believe that p if p is false.

On my analysis of truth tracking and the predicative analysis of SATs, SATs can't track the truth. This would require that $\Pr(B(T\ \iota x\ B_{p}x)|T\ \iota x\ B_{p}x)$ is substantially higher than $\Pr(B(T\ \iota x\ B_{p}x)|{\sim}T\ \iota x\ B_{p}x)$. But this condition cannot be met for the same reasons that we gave to show that, on the predicative analysis, SATs cannot be sensitive. On the Strawsonian construal of definite descriptions built into the predicative analysis of SATs, the probability that you believe $T\ \iota x\ B_{p}x$

if $T\,\iota x\, B_p x$ is false is the probability that you believe $T\,\iota x\, B_p x$ if p is false *and you believe p*. But so long as you believe p, the probability that you believe $T\,\iota x\, B_p x$ can be expected to be unaffected by the truth value of p. Hence the tracking ratio of a SAT will always be 1. Therefore, on my construal of truth tracking, SATs, on the predicative analysis, can't track the truth.

This wouldn't be a problem if we could have adequate evidence for SATs, since I contend that it is in principle possible to have adequate evidence for a proposition p, and hence to know it, even though your belief that p doesn't track the truth (Zalabardo 2012: 63-66). But I argue that this is not a viable solution to our problem, since it's not possible to have adequate evidence in support of SATs (Zalabardo 2012: 155-62). It follows that, on the predicative construal of SATs, my account of knowledge makes it impossible for SATs to be knowledge. My proposal (call it *SRB*) faces a version of the problem of reflective knowledge.[5]

On the assumption that the predicative analysis of SATs is correct, we have a powerful argument against SRB. However, the connection cuts both ways. To the extent that SRB is independently supported, we have at our disposal a powerful reason for abandoning the predicative analysis in favour of alternatives for which the problem doesn't arise. The argument would go like this:

(1) If SRB is the right account of knowledge and the predicative analysis of SATs is correct, then reflective knowledge is impossible.

(2) SRB is the right account of knowledge.

(3) Reflective knowledge is possible.

Therefore

(4) The predicative analysis of SATs is incorrect.

Call this the *fightback argument*. I think that premise (1) is incontestable, and premise (3) is highly plausible. I also believe premise

[5] Adam Leite has argued that the problem can be solved by thinking of SATs as standing beliefs (Leite 2014: 161). I don't think Leite's proposal is satisfactory. See Zalabardo 2014: 196.

(2). Hence I think this is a sound argument against the predicative analysis. But for the argument to work, there needs to be a plausible alternative analysis of SATs for which SRB doesn't face the problem of reflective knowledge. If this alternative analysis couldn't be found — if there were no plausible analyses of SATs for which SRB doesn't face the problem of reflective knowledge, then it would be hard to resist the conclusion that the problem rests, not with our analysis of SATs, but with my analysis of knowledge. My goal in the remainder is to fill this gap — to identify a plausible analysis of SATs for which SRB doesn't face the problem of reflective knowledge.

5 Deflationism

The most visible alternative to the predicative analysis is deflationism. According to deflationism, a SAT has the same content as its object of predication. The content of your belief that your belief that *p* is true is identical to the content of your belief that *p*. Contrary to what the surface grammar suggests, your belief that your belief that the cat is hungry is true is not about the instantiation of a property (truth) by one of your beliefs — it is about the instantiation of a property (hunger) by the cat. It has the same *cognitive content* as your belief that the cat is hungry.

The only difference between the two beliefs is that the former, but not the latter, carries an existential commitment to your belief that the cat is hungry. Adapting Hartry Field's terminology, we can characterise the resulting relationship between the two beliefs by saying that their cognitive equivalence is relative to the existence of your belief that the cat is hungry, where relative cognitive equivalence is to be understood as follows:

> To say that A is cognitively equivalent to B relative to C means that the conjunction of A and C is cognitively equivalent to the conjunction of B and C; so that as long as C is presupposed we can treat A and B as equivalent. (Field 1994: 250)

So long as it is presupposed that you believe that the cat is hungry, we can treat the two beliefs as equivalent. This gives to your belief that your belief that the cat is hungry is true the truth conditions that we expect from a SAT: true if you believe truly that the cat is hungry, false if you believe falsely that the cat is hungry, and neither true nor false if you don't believe that the cat is hungry, as in this case the existential commitment of your belief that your belief that the cat is hungry is true is not satisfied.[6]

On the analysis of SATs that we obtain from the deflationist account, the problem of reflective knowledge doesn't arise. The content of your belief that your belief that the cat is hungry is true is the same as the content of your belief that the cat is hungry. It follows that they both will have the same epistemic status. If you know that things are as your belief that the cat is hungry represents them as being, then you must also know that things are as your belief that your belief that the cat is hungry is true represents them as being, since the way the latter represents things as being is identical with the way the former represents things as being. We have one, not two, possible items of knowledge.

It seems then that deflationism provides what the fightback argument requires — an analysis of SATs on which SRB doesn't face the problem of reflective knowledge. This enables the advocate of SRB to blame the problem on a defective analysis of SATs, and save her account of knowledge by endorsing the deflationist alternative to the predicative analysis.

This is not a route I would like to take. Epistemological dividends notwithstanding, I believe that the deflationist account of truth faces serious obstacles. If it turned out that saving SRB from the problem of reflective knowledge required endorsing deflationism, I would be inclined to join others in concluding that the problem of reflective knowledge is a symptom of a defective epistemology. This is not the place to undertake a serious assessment of deflationism, but I want to indicate briefly in the next section the general source of my misgivings about the view.

6 Truth and meaning

My reservations arise from a consequence of deflationism first highlighted by Michael Dummett:

> It now appears that if we accept the redundancy theory of 'true' and

[6] See also Peter Strawson's analysis of truth ascriptions in Strawson 1949.

> 'false' […] we must abandon the idea which we naturally have that the notions of truth and falsity play an essential role in any account of the meaning of statements in general or of the meaning of a particular statement. (Dummett 1978: 7)

The "idea which we naturally have" is one that I find very plausible: in order to explain the power of statements or beliefs to represent things as being a certain way, we will need to invoke, as a crucial part of our explanans, the idea that statements and beliefs are made true or false by how things stand. In order to understand what it means for statements or beliefs to represent things as being a certain way, we *first* need to understand what it means to assess them according to whether the way they represent things as being agrees with the way things are — we need to understand, in other words, what it means to assess them as true or false. I am going to refer to assessment of beliefs, statements and similar items as true or false as *alethic assessment*. And I'm going to refer to the claim that our account of representation must be built on our account of alethic assessment as the *principle of alethic priority*. I am not going to defend alethic priority here. The principle will figure in my argument as a premise.[7]

Dummett's point is that alethic priority is incompatible with deflationism.[8] If our account of alethic assessment is going to contribute to our account of the representational features of statements or beliefs, then our account of alethic assessment can't invoke the representational features of these items. But this is precisely what deflationism does. Deflationism explains what it means to assess a belief as true in terms of the representational features of the belief. It explains the content of your belief that your belief that the cat is hungry is true in terms of (as identical with) the content of your belief that the cat is hungry. Hence the attempt to combine alethic priority with deflationism produces a vicious circle of explanation:

> in order that someone should gain from the explanation that P is true in such-and-such circumstances an understanding of the sense of P, he must already know what it means to say that P is true. If when he enquires into this he is told that the only explanation is that to say that P is

[7] For a recent attack on alethic priority, see Rumfitt 2014.

[8] See Collins 2002 for an interesting discussion of this point.

> true is the same as to assert P, it will follow that in order to understand what is meant by saying that P is true, he must already know the sense of asserting P, which was precisely what was supposed to be explained to him. (Dummett 1978: 7)

The incompatibility of deflationism with alethic priority is readily accepted by the leading deflationists. Their reaction is, understandably, to reject alethic priority. Thus, according to Hartry Field,

> [...] the main idea behind deflationism [...] requires [...] that what plays a central role in meaning and content not include truth conditions (or relations to propositions, where propositions are conceived as encapsulating truth conditions). (Field 1994: 253)[9]

Field uses verificationism as an illustration of the kind of account that might be used by the deflationist to explain the content of sentences. Horwich mentions assertibility conditions in this connection (Horwich 1990: 73), and, more recently, patterns of sentence acceptance that provide the causal-explanatory basis for our overall use of words (Horwich 2005: 49-50). What matters for our purposes is that all these proposals violate the principle of alethic priority, and that only accounts that violate the principle are compatible with deflationism. I want to uphold alethic priority. That's why I must reject deflationism.

Where does this leave us? We saw in Section 4 that the fightback argument requires that we identify a plausible alternative to the predicative analysis of SATs for which SRB doesn't face the problem of reflective knowledge. We then saw in Section 5 that on the deflationist analysis of SATs the problem of reflective knowledge does not arise. But in this section we've seen that if one wants to uphold alethic priority, as I do, endorsing deflationism is not an option. This still leaves us in need of an alternative to the predicative analysis to underpin the fightback argument. What we are looking for is an analysis of SATs that frees SRB from the problem of reflective knowledge without violating alethic priority.

[9] See also Horwich 1990: 71-74.

7 Explicating alethic assessment

We obtained the deflationist alternative to the predicative analysis from an account of alethic assessment — restricted to the first-person present case, i.e. of the content of beliefs in which you assess as true your own current beliefs. The deflationist explains the content of these assessments as identical with the content of the assessed beliefs. I am rejecting this approach on the grounds that it violates alethic priority. What we are after is an account of alethic assessment that abides by alethic priority but doesn't force us to adopt the predicative analysis of SATs.

A very natural strategy for explicating alethic assessment violates the second requirement. One way to explain the meaning of assessing *X*s as *Y*s is to specify what an *X* has to be like in order to qualify as a *Y* — the condition that an *X* has to satisfy in order for this assessment to be correct. The strategy I have in mind applies this general template to alethic assessment: we explain what it means to assess a belief as true by specifying the condition that a belief has to satisfy in order to count as true. I'm going to refer to this strategy for explicating alethic assessment as *representationalism*. The representationalist strategy can be implemented in many different ways. One prominent option is to think of belief as a relation to sentences of a language or language-like medium of representation, and to specify the condition that makes a belief true in terms of a Tarski-style theory of truth for these sentences, based on a theory of reference for their terms. But this is not by any means the only option available to the representationalist. An account of what makes a belief true in terms of, say, coherence, or end-of-enquiry consensus, would also sustain a representationalist account of alethic assessment.[10]

[10] It's an interesting question whether a deflationist account of truth could also give rise to a version of representationalism. I think the question has to be answered in the negative if deflationism satisfies the condition that Field imposes on the view if it is to be at all interesting: "it must claim not merely that what plays a central role in meaning and content not include truth conditions *under that description, but that it not include anything that could possibly constitute a reduction of truth conditions to other more physicalistic terms*" (Field 1994: 253). However, it seems to me that Paul Horwich's account of meaning doesn't satisfy this condition, and can give rise to a representationalist explication of alethic assessment. See Horwich 2005, especially Chapter 2.

The problem with any kind of representationalism, for our purposes, is that it enjoins the predicative analysis of SATs. If we explicate alethic assessment by specifying the condition that a belief has to satisfy in order to count as true, then a SAT can only be a belief to the effect that this condition is satisfied by one of your current beliefs, as the predicative analysis dictates. If we want to find an alternative to the predicative analysis of SATs, we need to adopt a non-representationalist strategy for explicating alethic assessment.

In the remainder I'm going to explore an alternative to the representationalist strategy for which I'm going to use the label *pragmatism*. The pragmatist rejects the representationalist project of explicating alethic assessment with a specification of the condition that a belief has to satisfy in order to be assessed as true. What the pragmatist proposes instead is to render alethic assessments intelligible with a specification of the rules that govern the practice of assessing beliefs in this way. For the pragmatist, alethic assessment is assessment that follows these rules, and 'true' is the label that we apply to beliefs or other items in order to express a favourable assessment according to these rules. In the next section I'm going to outline a proposal as to which rules we should take to define alethic assessment. I'm going to refer to this specific pragmatist account of alethic assessment as *empiricist pragmatism*.[11]

8 Empiricist pragmatism[12]

I'm going to concentrate in the first instance on alethic assessment, not of beliefs, but of the episodes that we think of as conscious manifestations of belief — conscious episodes in which we take ourselves to represent things as being a certain way.[13] I'm going to refer to these episodes as *judgments*, although the term sometimes carries a connotation of voluntariness or spontaneity that will be absent from

[11] The label is meant to mark the contrast with Robert Brandom's rationalist pragmatism (Brandom 2000: 11). For the contrast see Zalabardo 2016, Section 5.

[12] This section overlaps with Zalabardo 2016.

[13] Brandom's rationalist pragmatism focuses in the first instance on assertion, as the linguistic correlate of judgment (Brandom 1994: 153). Brandom's reasons for taking this line do not apply to my proposal.

my account. My characterisation of the rules that govern alethic assessment of judgments will rest on some substantial assumptions about the nature of judgments. But since we want the resulting account to abide by the principle of alethic priority, our assumptions regarding the phenomenon cannot include semantic features — the fact that they represent things as being a certain way.

I want to take as my starting point David Hume's characterisation of the episodes that I'm calling judgments, but he identifies with beliefs, in the Appendix to the *Treatise*. He writes:

> belief consists merely in a certain feeling or sentiment; in something, that depends not on the will, but must arise from certain determinate causes and principles, of which we are not masters. When we are convinc'd of any matter of fact, we do nothing but conceive it, along with a certain feeling, different from what attends the mere *reveries* of the imagination. (Hume 1978: 624)

Belief, according to Hume, then, is a conscious involuntary reaction. What it is a reaction to is not, in the first instance, the possible state of affairs that the belief represents as obtaining, but the idea that serves as its representative in the mind:

> an opinion or belief is nothing but an idea, that is different from a fiction [...] in the *manner* of its being conceiv'd. (Hume 1978: 628)

> An idea assented to *feels* different from a fictitious idea, that the fancy alone presents to us. (Hume 1978: 629)[14]

I want to focus on the phenomenon that Hume highlights, not as an account of belief, but as the basis for a characterisation of the kind of conscious episodes that I'm calling *judgments*. Judgments will have the basic character that Hume ascribes to beliefs — they are conscious episodes in which a mental item produces an involuntary reaction.

I am going to use the term *conviction* for the conscious, involuntary, re-identifiable reaction (Hume's feeling or sentiment) that figures in judgments. I'm going to complicate Hume's picture slightly

[14] I am not adopting Hume's account of the difference between these episodes and those in which a possible state of affairs is merely imagined, in terms of "a superior *force, or vivacity, or solidity, or firmness, or steadiness*" (Hume 1978: 629).

by contemplating *negative conviction*, as the feeling associated with things not being as represented in consciousness, as well as *positive conviction*. I will refer to judgments as either positive or negative, depending on the sign of the conviction that figures in them. I want to emphasize that I'm thinking of conviction as a *feeling*. Conviction doesn't ascribe a property or concept to a possible state of affairs or to its mental representative,[15] nor is it the undertaking of a commitment of any kind. It is simply an involuntary feeling that some conscious items provoke.[16] Conceiving of conviction along these lines doesn't require assuming that it has a particularly rich phenomenology. There doesn't have to be a collection of phenomenological features that are present precisely in those conscious episodes that involve conviction. All that's required is that the subject has the ability to re-identify this feeling. Its type-identity conditions can then be defined in terms of the subject's verdicts.

To the conscious items that judgments are reactions to, I am going to refer as *conscious sentences*. They will be the representatives in the stream of consciousness of the possible states of affairs that we take judgments to represent as obtaining, leaving out of the picture for now the possible semantic properties of these mental entities. Like the sentences of a natural or formal language, they exhibit syntactic, combinatorial structure, being produced by the combination of constituents (*conscious terms*) according to specific patterns. Like Hume's ideas, conscious sentences will figure in conscious episodes other than judgments, including the conscious, episodic correlates of desire (the kind of conscious episode that occurs, for example, when you obey the order to close your eyes and make a wish) or episodes in which we merely consider in consciousness a way for things to be, without taking any attitude towards it.[17]

[15] Hume considers and rejects this option, as the view that "belief is some new idea, such as that of *reality or existence, which we join to the simple conception of an object*" (Hume 1978: 623).

[16] See in this connection Horgan and Timmons' discussion of the phenomenological dimension of what they call *occurrent beliefs* in Horgan and Timmons 2006. See also Jonathan Cohen notion of *credal feelings* (Cohen 1992).

[17] Notice that what I am calling conscious sentences are importantly different from the sentences of the language of thought postulated by the represen-

Conscious sentences may appear spontaneously in the stream of consciousness, or they might be produced voluntarily. When a conscious sentence figures in the stream of consciousness, we may feel towards it positive conviction, negative conviction, or neither.[18] Which of these obtains in each case is not under the control of the will, but, as Hume indicates, it's not a random matter either — conviction arises "from certain determinate causes and principles". To judge, on my pre-semantic construal, is simply to feel conviction towards a conscious sentence.

I have characterised conscious sentences as certain re-identifiable items that can be brought to consciousness voluntarily or appear there spontaneously, and conviction as a specific involuntary reaction that we may or may not feel towards a conscious sentence that we are entertaining. Judgments are the episodes in which this reaction is produced. We think of conscious sentences and judgments as representing things as being a certain way, but our characterisation of these phenomena doesn't presuppose that they have this power. Hence by invoking this characterisation of judgments in our account of alethic assessment we won't be violating alethic priority.

A pragmatist account of alethic assessment of conscious sentences and the judgments in which they figure will proceed by specifying a collection of rules such that an assessment practice will count as alethic just in case it is governed by these rules. According to empiricist pragmatism, the practice of alethic assessment is governed by three rules: the Basic Rule, the Ascent Rule and the Interpretation Rule.

According to the Basic Rule, alethic assessment is necessarily driven by conviction. To assess conscious sentences in any other way is not to assess them as true or false:

tational theory of mind. Conscious sentences, unlike sentences of the language of thought, are essentially conscious, enjoying no ontological status beyond the conscious episodes in which they figure.

[18] Conviction comes in degrees, and the phenomenon might be more accurately represented as a continuum between 1 and 0, with .5 as the complete absence of positive or negative conviction. However, I'm going to proceed, for the sake of simplicity, as if there were sharp boundaries between the presence of each type of conviction and their absence.

> *Basic Rule*: Assess a conscious sentence as true if and only if it produces positive conviction; assess a conscious sentence as false if and only if it produces negative conviction.

Notice the parallel with some expressivist accounts of specific regions of discourse. According to a version of expressivism concerning moral discourse, to assess an action as morally right or wrong is to assess it according to your moral sentiments — to assess it as morally right when it produces moral approval in you and as morally wrong when it produces moral disapproval.[19]

Clearly the basic rule by itself doesn't provide a sufficient characterisation of alethic assessment. One major limitation is that it is compatible with a highly implausible subjectivism, as it makes no provision for treating as incorrect a judgment that follows the subject's convictions. We can see this in the first instance with respect to one's past judgments. A subject can presumably entertain a conscious sentence on two different occasions, and it is perfectly possible that it produces conviction on one occasion but not on the other, or that it produces positive conviction on one occasion and negative conviction on the other. This might happen as a result of changes either in the subject's state of information or in the processes that determine the production of conviction in her.

The Basic Rule gives no grounds for treating judgments of opposite signs concerning a single conscious sentence as incompatible with one another, or one's previous judgments as false. The Basic Rule by itself would confer on alethic assessment the behaviour of forms of assessment for which a subjectivist construal is perfectly adequate. Consider, for example, the plausible view that to assess an ice-cream flavour as delicious or revolting is to assess it according to your culinary taste — as delicious if it gives you gustatory pleasure and as revolting if it gives you gustatory displeasure. Tastes change and you might find that if you follow this rule you end up assessing pistachio ice-cream as revolting on one occasion and as delicious a

[19] The claim that I'm focusing on is that assessment of actions has to be conducted in this way in order to count as moral assessment, not the claim that the role of moral discourse is to express moral sentiments or a claim to the effect that a moral assessment is correct just in case it accords with the moral sentiments of the assessor.

few years later. There is no obvious sense in which these assessments are in conflict with one another. If the Basic Rule were the only rule governing alethic assessment, we'd have to treat in the same way the situation in which a subject goes from assessing a conscious sentence as true to assessing it as false.

In order to address this issue, we need to introduce a rule that enables us to go from assessments of conscious sentences to assessments of judgments:

> *Ascent Rule*: Assess a positive judgment of a conscious sentence as true and a negative judgment of the sentence as false if and only if you assess the conscious sentence as true; assess a positive judgment of a conscious sentence as false and a negative judgment of the sentence as true if and only if you assess the conscious sentence as false.[20]

In order to abide by this rule, a subject who now feels negative conviction towards a conscious sentence but remembers feeling positive conviction towards the same sentence in the past will also have to assess as false her past judgment. The same would go for a subject who now feels positive conviction towards a conscious sentence but remembers feeling negative conviction towards it.

Notice that this feature of alethic assessment resembles a parallel feature of moral assessment. When we assess an action as morally right, we also assess as morally right moral approval of the action and we assess as morally wrong moral disapproval of it. Likewise, when we assess an action as morally wrong, we also assess as morally right moral disapproval of the action and we assess as morally wrong moral approval of it.

The practice described by the Basic Rule and the Ascent Rule still has a very important limitation — it imposes no restrictions on how I should assess the judgments of others. The limitation wouldn't exist if we could make sense of the idea that one of your conscious sentences is identical to one of mine, but it is hard to see how this could be achieved. For a single subject, we can think of the identity

[20] This formulation of the rule presupposes that the sentences in question have no indexical features. Dealing with indexicality would require a more sophisticated approach. The same goes for the next rule.

conditions of conscious sentences as given by the subject's inclinations — two conscious episodes involve the same conscious sentence just in case it seems to the subject that they do. For inter-personal identity there is no obvious correlate for this approach.

A plausible account of the rules that govern alethic assessment would have to impose conditions on our assessment of the judgments of others. It is an essential feature of the practice that we can assess as true or false the judgments of others, and there are some conditions that these third-person assessments have to satisfy in order to count as alethic. The basic intuitive idea of the rule we need is very simple: in order for your assessment to count as alethic assessment, you need to assess as true those judgments of others that agree with yours, and you need to assess as false those judgments of others that disagree with yours.

Unfortunately, however, the rule cannot be formulated in these simple terms, as invoking at this point the idea of someone else's judgment agreeing or disagreeing with one of yours would render the account incompatible with the principle of alethic priority. In order for your judgment to agree or disagree with mine, the way things are represented as being by the conscious sentence that produces your conviction has to coincide with the way things are represented as being by the conscious sentence that produces mine. Availing ourselves of this notion would amount to invoking semantic features of judgments in our account of alethic assessment.

The way forward for the pragmatist at this point is to invoke the phenomenon of interpretation.[21] We have introduced conscious sentences as the representatives in the stream of consciousness of possible states of affairs and the immediate objects of conviction. But conscious sentences play an important additional role: we use them to index or tag the representational states of others, including their judgments and the beliefs they manifest, in the procedure that we

[21] I think there are important similarities between the role that interpretation plays in this construal of alethic assessment and the role that it plays, according to Donald Davidson, in the concept of truth (Davidson 1990: 295-96). John Collins offers an insightful summary of Davidson's line of reasoning on this point (Collins 2002: 520). For a defence of the pragmatist character of Davidson's position, see Rorty 1986.

refer to as *interpretation*.[22] We can think of these indexings as conscious sentences that embed other conscious sentences, postulating a relation between our interpretee's judgment and the embedded conscious sentence. These interpretative conscious sentences, like our other conscious sentences, may or may not produce conviction, positive or negative, when they are brought to consciousness.

The judgments that we interpret as agreeing with ours are those that we index with conscious sentences towards which we feel conviction of the same sign (positive or negative); the ones that we interpret as disagreeing with ours are those that we index with conscious sentences towards which we feel conviction of the opposite sign.[23] This feature of the practice is represented in our final rule:

> *Interpretation Rule*: Assess someone else's positive judgment as true and someone else's negative judgment as false if and only if you have indexed it with a conscious sentence that produces positive conviction in you. Assess someone else's positive judgment as false and someone else's negative judgment as true if and only if you have indexed it with a conscious sentence that produces negative conviction in you.[24]

9 Empiricist pragmatism and the fightback argument

Empiricist pragmatism proposes to take the Basic Rule, the Ascent Rule and the Interpretation Rule as defining the practice of alethic

[22] For the picture of interpretation as an indexing exercise, see Churchland 1979: 100-7. In Churchland's version of the approach, the items that serve as indices are propositions, but he sees viability as independent of any special view concerning the nature of propositions. He thinks the approach would work even if we thought of propositions as sentences.

[23] Huw Price has highlighted the need for a rule along these lines in the characterisation of our conversational practice (Price 2011: 164).

[24] This formulation of the rule would still be at odds with alethic priority if interpretation were defined in terms of the goal of matching the judgments of others with conscious sentences of yours that are synonymous with the conscious sentences that serve as the objects of those judgments. The proposal needs to employ a construal of interpretation on which its goal is not defined in semantic terms. See Zalabardo 2016 for further discussion.

assessment. To assess judgments as true or false is, the pragmatist claims, to assess them according to these rules.[25] The approach satisfies our criteria. Notice, first, that it sustains an analysis of SATs that doesn't pose the problem of reflective knowledge. Alethic assessment of your own current judgments necessarily follows the convictions that figure in these judgments — positive judgments can only be assessed as true; negative judgments can only be assessed as false. Assessments that depart from this pattern simply don't qualify as alethic. It follows that if we think of alethic assessments as judgments, then your alethic assessment of one of your current judgments (i.e. a SAT) will have to be treated as having the same content as the target judgment — the way the assessment represents things as being will have to coincide with the way the target judgment represents things as being. Hence, if you know that things stand as the target judgment represents them, there is no further question of whether you know that they stand as your alethic assessment — your SAT — represents them. The problem of reflective knowledge doesn't arise. But, second, this is achieved without violating alethic priority, since the pragmatist explication of alethic assessment doesn't invoke semantic features of judgments. Truth, as construed by empiricist pragmatism, can then figure in our account of the content of judgments without circularity.

We've seen that the advocate of SRB could try to defuse the threat of the problem of reflective knowledge with the fightback argument, arguing that the problem arises from an inadequate analysis of SATs. But pursuing this strategy requires identifying an alternative to the predicative analysis of SATs for which the problem doesn't arise. Deflationism can play this role so long as we are prepared to forsake alethic priority. Empiricist pragmatism serves the same purpose without paying this price. For the advocate of SRB (or SC) and alethic priority, the problem of reflective knowledge provides support for empiricist pragmatism.[26]

[25] See Zalabardo 2016 for some ways in which the proposal will need to be fine-tuned.

[26] I am grateful to Josep Corbí and Jordi Valor for inviting me to take part in the 1st Blasco Disputatio. The present paper is not so much my contribution to the symposium as an attempt to address concerns raised by my co-symposiast, Marian David.

José L. Zalabardo
Philosophy Department
UCL
Gower Street
London WC1E 6BT
UK
j.zalabardo@ucl.ac.uk

References

Becker, Kelly. 2006. Is counterfactual reliabilism compatible with higher-level knowledge? *Dialectica* 60 (1): 79-84.

Brandom, Robert. 1994. *Making It Explicit*. Cambridge, MA: Harvard University Press.

Brandom, Robert. 2000. *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, MA: Harvard University Press.

Churchland, Paul M. 1979. *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press.

Cohen, L. Jonathan. 1992. *An Essay on Belief and Acceptance*. Oxford: Clarendon.

Collins, John. 2002. Truth or meaning? A question of priority. *Philosophy and Phenomenological Research* 65: 497-536.

Davidson, Donald. 1990. The structure and content of truth. *Journal of Philosophy* 87: 279-328.

DeRose, Keith. 1995. Solving the skeptical problem. *Philosophical Review* 104: 1-52.

Dummett, Michael. 1978. Truth. In *Truth and Other Enigmas*. London: Duckworth.

Field, Hartry. 1994. Deflationist views of meaning and content. *Mind* 103: 249-85.

Horgan, Terry, and Mark Timmons. 2006. Cognitive expressivism. In *Metaethics After Moore*, edited by T. Horgan and M. Timmons. Oxford: Oxford University Press.

Horwich, Paul. 1990. *Truth*. Oxford: Basil Blackwell.

Horwich, Paul. 2005. *Reflections on Meaning*. Oxford: Oxford University Press.

Hume, David. 1978. *A Treatise of Human Nature*. 2nd ed. Oxford: Clarendon.

Leite, Adam. 2014. Standing beliefs, skepticism, and some questions about Zalabardo's probabilistic approach. *Teorema* 33: 159-70.

Melchior, Guido. 2015. The heterogeneity problem for sensitivity accounts. *Episteme* 12: 479-96.

Price, Huw. 2011. Truth as convenient friction. In *Naturalism Without Mirrors*.

This research has been supported by the Ministerio de Economia y Competitividad of the Government of Spain (project FFI2012-38908-C02-01). This support is gratefully acknowledged.

Oxford: Oxford University Press.

Rorty, Richard. 1986. Pragmatism, Davidson, and truth. In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, edited by E. LePore. Oxford: Blackwell.

Roush, Sherrilyn. 2005. *Tracking Truth*. Oxford: Oxford University Press.

Rumfitt, Ian. 2014. Truth and meaning. *Aristotelian Society Supplementary Volume* 88: 21-55.

Salerno, Joseph. 2010. Truth-tracking and the problem of reflective knowledge. In *Knowledge and Skepticism*, edited by J. Keim Campbell, M. O'Rourke and H. S. Silverstein. Cambridge, Mass.: MIT Press.

Sosa, Ernest. 1999. How to defeat opposition to Moore. In *Philosophical Perspectives, 13, Epistemology*, edited by J. E. Tomberlin. Malden, Massachusetts and Oxford: Blackwell.

Strawson, Peter F. 1949. Truth. *Analysis* 9: 83-97.

Strawson, Peter F. 1950. On referring. *Mind* 59: 320-34.

Vogel, Jonathan. 1987. Tracking, closure, and inductive knowledge. In *The Possibility of Knowledge: Nozick and His Critics*, edited by S. Luper-Foy. Totowa, New Jersey: Rowman & Littlefield.

Vogel, Jonathan. 2000. Reliabilism leveled. *Journal of Philosophy* 97: 602-23.

Zalabardo, José L. 2012. *Scepticism and Reliable Belief*. Oxford: Oxford University Press.

Zalabardo, José L. 2014. Replies to my critics. *Teorema* 33: 181-202.

Zalabardo, José L. 2016. Empiricist pragmatism. *Philosophical Issues* 26.