PROBABILIST ANTIREALISM

BY

IGOR DOUVEN, LEON HORSTEN AND
JAN-WILLEM ROMEIJN

Abstract: Until now, antirealists have offered sketches of a theory of truth, at best. In this paper, we present a probabilist account of antirealist truth in some formal detail, and we assess its ability to deal with the problems that are standardly taken to beset antirealism.

1. Introduction

According to antirealists, there is an intimate connection between truth and human cognitive capacities that holds by conceptual necessity. While antirealists differ about the exact nature of the connection, no antirealist disputes its conceptual necessity; they distinguish the antirealist conception of truth from a realist one accompanied by some methodological view to the effect that, by natural selection perhaps, or maybe just by good fortune, our epistemic powers happen to be so attuned to the world we inhabit that there exist no truths which are beyond our ken in principle. So far, antirealists have proposed constraints to be met by antirealist theories of truth, and even a sporadic "informal elucidation" of antirealist truth (Putnam, 1981, p. 56), but an antirealist *theory* of truth, comparable, if only just remotely, in formal precision to Tarski's (1956) theory of truth, for instance, is still glaringly missing from the literature. Williamson (2006) is right to castigate antirealists for, so far at least, failing to offer anything going beyond a merely programmatic sketch of their position. The present paper aims to address this lacuna by stating a formally precise probabilist account of antirealist truth for a language. We should like to stress that no attempt will be made in the following to convert anyone to antirealism. Rather, we wish to put forth a formal theory of truth that should appear attractive to those who are already attracted by the antirealist idea that truth is an inherently epistemic notion.

Pacific Philosophical Quarterly 91 (2010) 38–63

The adequacy conditions for a theory of antirealist truth are partly the same as those for a theory of realist truth. The theory should be both materially and formally adequate in Tarski's sense. That is, the truth predicate, as defined by the theory, should satisfy the disquotational schema and the theory should be paradox-free. In addition, it should not entail what one might call quasi-paradoxes, that is, consistent but intuitively absurd claims, such as – to mention a famous example – the claim that all truths are known. Furthermore, the theory should be intuitively correct in that it should make most, and preferably all, sentences anybody would unproblematically regard as being truth-valued come out as such. Likewise, it should validate certain platitudes about truth, such as that a conjunction is true if and only if both of its conjuncts are true. And, of course, if the theory is to offer a definition of *antirealist* truth, it should secure a conceptual tie between truth and the epistemic. In fact, the tie should be such as to render the theory responsive to the considerations that have tended to motivate antirealists.

We begin, in Section 2, by stating the core of our probabilist theory and by addressing some concerns that one might have about it. Then, in Sections 3–6, we consider how the theory fares with respect to the aforementioned adequacy conditions and argue that, on the whole, and as far as our assumptions allow us to conclude, the theory does well on this count. Finally, we show that the theory compares favorably with Putnam's informal elucidation of antirealist truth, and this not merely on the count of formal precision (Section 7).

2. *Antirealist truth defined*

The theory we are about to offer can be regarded as a formalization of the Peircean view of truth, which equates truth with "[t]he opinion which is fated to be ultimately agreed by all who investigate" (Peirce, 1978, 5.407; see also Peirce, 1978, 5.565). We aim to make this idea precise for a given language by employing the machinery of Bayesian epistemology. We start by making some assumptions about the language and by briefly rehearsing the central Bayesian tenets.

2.1 THE LANGUAGE

We are giving a definition of truth for a language $\mathcal{L}$ which we suppose to be a regimented language in which empirical scientific theories can be expressed.

$\mathcal{L}$ is a first-order language. Its vocabulary includes the usual logical vocabulary. It also includes mathematical vocabulary, and some non-mathematical vocabulary; it has individual constants d_0, d_1, d_2, ... We need not be precise about exactly which mathematical and non-mathematical constants and predicates are included. But at the outset, we do not include the truth predicate Tr; this is considered to be a metalinguistic notion. And since we plan to reduce truth to degrees of belief or subjective probabilities, the (subjective) probability operator is also considered as a metalinguistic notion. We think of $\mathcal{L}$ as an *interpreted* language and assume that the domain of every model for $\mathcal{L}$ is either finite or denumerably infinite, and that every object of the domain is named by an individual constant. Whenever we speak of the sentences of $\mathcal{L}$, we mean the *declarative* sentences of the language (or *statements*, as some would say). Lower-case Greek letters serve both as linguistic and as metalinguistic sentence variables; we trust that context will suffice to distinguish between the two uses.

We further assume that there is a designated part $\mathcal{E} \subset \mathcal{L}$ of the language such that all and only sentences belonging to $\mathcal{E}$ are apt to report evidence. Sentences that are not evidence sentences are called "theoretical sentences"; $\mathcal{T} = \mathcal{L} \setminus \mathcal{E}$ is the class of theoretical sentences. Exactly how the two classes are to be delineated will not detain us here. One could perhaps characterize the evidence sentences as those sentences that rational agents are willing to assign probability 1 as a direct effect of experience, where the other sentences in $\mathcal{L}$ can have their probability altered only mediately, because some evidence sentence receives probability 1. Alternatively, one might try to define evidence sentences syntactically, for instance, as the atomic sentences whose predicate, function and constant symbols all belong to the "observational" part of the vocabulary. In the following, however, we rely on an intuitive understanding of the notion of evidence sentence, as is in effect customary among Bayesian epistemologists.

Finally, we assume that $\mathcal{L}$ is governed by classical logic. Although this is not the preferred choice of logic of all who call themselves antirealists, it is certainly not antithetical to antirealism either. For instance, Peirce and (middle) Putnam, who unambiguously qualify as antirealists in the present sense, both accept classical logic.

2.2 PROBABILITY

We assume a community of rational agents, roughly corresponding to the Peircean community of "all who investigate." An agent is supposed to have a degrees-of-belief function defined on all sentences of $\mathcal{L}$, and he or she is said to be rational iff he or she satisfies the following three conditions: first, his or her degrees of belief at all times are representable by a probability function, where a probability function is a function Pr satisfying these axioms:

(A1) $0 \leqslant \Pr(\varphi) \leqslant 1$ for all $\varphi \in \mathcal{L}$;
(A2) $\Pr(\varphi) = 1$ for all $\varphi \in \mathcal{L}$ such that φ is a logical truth;
(A3) $\Pr(\varphi \vee \psi) = \Pr(\varphi) + \Pr(\psi)$ for all $\varphi, \psi \in \mathcal{L}$ such that φ is logically inconsistent with ψ;
(A4) $\Pr(\exists x \varphi(x)) = \lim_{n\to\infty} \Pr\left(\bigvee_{i=0}^{n} \varphi(d_i)\right)$ for all open formulas $\varphi(x) \in \mathcal{L}$ with at most x free.[1]

Second, his or her initial probabilities are strictly coherent. That is, before he or she has obtained any evidence, he or she assigns probability1only to logical truths and thus probability 0 only to logical falsehoods.[2] And third, as he or she receives new evidence, he or she updates his or her probabilities by dint of Bayes's rule. That is, for any given sentence φ, the agent's new probability for φ after he or she has become certain of ψ equals his or her earlier probability for φ conditional on ψ, where this is standardly defined to equal the probability of the conjunction of φ and ψ divided by the probability of ψ (provided the latter is greater than 0; else the conditional probability is undefined).

Strict coherence has been defended as a *general* requirement of rationality by various authors.[3] As such, it is problematic, however, given that strict coherence is incompatible with learning by means of Bayes's rule (which applies on the condition that one has become certain of a sentence one previously was uncertain of). Since we only require strictly coherent *initial* probabilities, there is no inconsistency in our definition. The requirement itself seems hardly more than common sense: how could one rationally assign extreme probabilities to empirical sentences before one has started to gather information about the world?[4]

We are going to define truth in terms of (subjective) probability. One might worry about a possible circularity of such a theory, for is "probability" not "probability of truth"? It should be remembered, however, that probability can be, and still standardly is, operationally defined in terms of betting dispositions.[5] Succinctly, one's probability for φ can be interpreted as the maximum price one is willing to pay for a bet on that sentence which pays \$1 if φ, and nothing otherwise. Naturally, there is nothing wrong with saying instead: ". . . which pays \$1 if φ *is true*, and nothing otherwise," given the disquotational schema $\varphi \leftrightarrow \mathrm{Tr}(\varphi)$, which Tr satisfies (at least in a strong enough sense), as will be seen in Section 3. But the use of the truth predicate is clearly dispensable here. For those who have qualms about the operationalist definition of probabilities, let us add that the foregoing is not to suggest that we are committed to that definition. If, for instance, subjective probabilities can be identified with brain states, which are measurable by a "psychogalvanometer" perhaps (as Ramsey (1926, p. 161) thought was at least conceivable), then the truth predicate may be equally dispensable for a proper definition of the notion of probability.

2.3 TRUTH FOR EVIDENCE SENTENCES

The definition of truth consists of two parts, one that defines truth for the elements of $\mathcal{E}$, and one that defines truth for the rest of $\mathcal{L}$. Let the non-empty set I be the community of rational agents. Then a very simple truth definition for $\mathcal{E}$ would state that an evidence sentence is true iff there is an $i \in I$ and a time t such that at t agent i assigns probability 1 to the sentence. However, this leads to inconsistency unless we assume that, either because of how $\mathcal{E}$ is delineated or because of the cognitive powers of rational agents (or because of a combination of the two), it can be excluded that, for some evidence sentence, some agents i and j, and some times t and t', i at t assigns probability 1 to the sentence while j at t' assigns probability 1 to its negation. We would thus seem to end up either with a very narrow class of evidence sentences – like, perhaps, sense data statements – or with very unrealistic idealizing assumptions about rational agents, which would leave little of the guiding antirealist thought that truth is intimately connected to *our* cognitive capacities. Of course, we could try other combinations of quantifiers in the definition, like "an evidence sentence is true iff for all agents/most agents/the majority of agents, there is a time at which they will assign probability 1 to it" or ". . . there is a time such that all (or most, or the majority of) agents. . ." But any of these combinations would still seem to result in a quite anemic theory of truth, making far too many sentences that pretheoretically have a truth value come out as lacking one, and thereby quite immediately failing to satisfy one of the earlier-mentioned adequacy conditions.

The following, subjunctive truth definition for elements of $\mathcal{E}$, which we recommend instead, does not share this defect:

$$\forall \varphi \in \mathcal{E}\left[\mathrm{Tr}(\varphi) \leftrightarrow \left\{\begin{array}{l}\text{for any } i \in I \text{ and any time } t, \text{ if } i \text{ were at } t \text{ in} \\ \text{circumstances sufficiently good for the appraisal} \\ \text{of } \varphi, \text{ then } i \text{ would at } t \text{ assign probability 1 to } \varphi\end{array}\right\}\right] \tag{1}$$

An evidence sentence that is not true is said to be false. As a result, all evidence sentences have a determinate truth value.

It merits remark that one could consider altering the first or second quantifier (or both) in the right-hand side of (1) to "for most . . . ," for instance, to allow for the occasional cognitive mishap that even rational agents may be expected to experience, even in circumstances being classified as "sufficiently good" for the appraisal of this or that sentence, without having to be overly restrictive in our choice of $\mathcal{E}$.

One possible worry about (1) is that we may not be able to define "sufficiently good conditions for the appraisal of φ" in any other way than as those conditions under which we can determine whether φ is *true*, thereby making the definition circular. The worry seems misplaced, however. To use an example of Putnam (1990), who proposed something very similar to (1) as applying to *all* sentences in the language (more on this in Section 7), sufficiently good circumstances for the appraisal of the sentence "There is a chair in my study" would be "to be in my study, with the lights on or with daylight streaming through the window, with nothing wrong with my eyesight, with an unconfused mind, without having taken drugs or being subjected to hypnosis, and so forth, and to look and see if there is a chair there" (p. viii). Clearly, there is no explicit appeal to the notion of truth here, and we submit (as no doubt Putnam does) that the "and so forth" could be spelled out in a way which does not make such an appeal either. But to give an *example* of how sufficiently good conditions can be specified without appealing to the notion of truth is not enough if (1) is supposed to be part of a definition of truth, for the latter would seem to require a prior definition of the notion of sufficiently good conditions for the appraisal of any given sentence. We think that, at least for evidence sentences, the hope is justified that such a definition can be had. If we assume that evidence sentences are what many philosophers of science, and certainly most philosophers engaged in the scientific realism debate, take them to be – namely, sentences attributing observable properties to observable entities or processes, or tuples of such entities or processes – then the conditions Putnam mentioned seem to already apply for many evidence sentences: that one's senses and mind be in good order, that there be enough light, that one be in relatively close proximity to the object(s) or process(es) the sentence is about, and that nothing obstruct one's view of the object(s) or process(es). Doubtless this will not do for all evidence sentences; observation is not always a matter of seeing, or not only a matter of seeing, but sometimes (also) of hearing or smelling or feeling. And, for instance, sufficiently good conditions for the appraisal of "My computer makes a humming sound," which by the aforementioned criterion would certainly seem to count as an evidence sentence, would include that it is (relatively) quiet in the room where the computer stands.[6] But this at most suggests the need for a definition with multiple clauses; for instance, one for sentences attributing a visual property to an observable entity or process, others for sentences attributing an auditory or an olfactory or a tactile property, and more besides perhaps. In any event, there seems to be no reason in principle to believe that the notion of sufficiently good conditions for the appraisal of evidence sentences cannot be generally characterized in a non-circular manner.

Another possible worry is that (1) might involve us in instances of Shope's (1978) conditional fallacy in that a sentence may have a truth value in the closest circumstances good enough for its appraisal that is different from the truth value it actually has. Timothy Williamson (in personal communication) pointed us to the sentences "Mary is blushing" and "I am being blown up by a bomb" as constituting potentially problematic examples in this regard. The worry, specifically, is that while Mary is not blushing, she would be, were we observing her, respectively, that the referent of the first-person pronoun in the second sentence is not being blown up by a bomb in any circumstances sufficiently good for appraising that sentence even if actually he or she *is* being blown up by a bomb. But, as for the first example, it seems that if the closest circumstances that are good enough for appraising whether Mary is blushing are ones in which we not only observe her but in which she also observes us, then the closeness or similarity relation between possible circumstances may well have been wrongly specified. And the second example may just give additional grounds for taking seriously the option of replacing "for any $i \in I$" in (1) by "for most $i \in I$."[7]

On a more fundamental level, it seems that the antirealist should dismiss these worries as being fueled by realist intuitions. At the root of the antirealist's thinking lies precisely the idea that it makes no sense to suppose that sentences may have truth values independently of our capacity to recognize them. *A fortiori*, it makes no sense to suppose that a sentence may actually have a truth value that differs from the one it has in the closest circumstances good enough for its appraisal.

2.4 TRUTH FOR ATOMIC THEORETICAL SENTENCES

To extend the above truth definition for evidence sentences to a truth definition for the entire language $\mathcal{L}$, we first define truth for atomic theoretical sentences.

Let $\mathcal{E}_{\mathrm{Tr}} \subset \mathcal{E}$ be the set of evidence sentences that are true according to (1), and let $A_t \subseteq \mathcal{E}_{\mathrm{Tr}}$ be the set of evidence sentences that are accepted by the community of rational agents at stage of inquiry t, meaning that at t all agents assign perfect probability to these sentences. It is assumed that at any given stage of inquiry there are only finitely many evidence sentences accepted by this community, so that A_t is finite for all t; "$\wedge A_t$" designates the conjunction of the elements of A_t. Further, let the sequence $\langle A_0, A_1, A_2, \ldots\rangle$ satisfy the conditions that $A_0 = \emptyset$ and $A_t \subset A_{t+1}$, for all t. Finally, Pr_i is agent i's probability function. Then truth for atomic sentences in $\mathcal{T}$ is defined thus:

$$\forall \text{ atomic } \varphi \in \mathcal{T}\left[\mathrm{Tr}(\varphi) \leftrightarrow \forall i \in I \lim_{t\to\infty} \mathrm{Pr}_i(\varphi|\wedge A_t) = 1\right], \tag{2}$$

and falsity for atomic sentences in $\mathcal{T}$ is defined thus:

$$\forall \text{ atomic } \varphi \in \mathcal{T}\left[\mathrm{F}(\varphi) \leftrightarrow \forall i \in I \lim_{t\to\infty} \mathrm{Pr}_i(\varphi|\wedge A_t) = 0\right]. \tag{3}$$

More than these two clauses is needed, unless we want to preempt the question of whether, for all atomic theoretical sentences, the relevant conditional probability assigned to it by any rational agent i will go either to 1 or to 0 “in the limit.” (Later on in the paper, we point to fairly general conditions which may preclude radical disagreement between agents in the sense that one rational agent assigns probability 0 to a sentence while another assigns probability 1 to that sentence.) If for some atomic sentence in $\mathcal{T}$ convergence in the sense specified here does not occur, we say that its truth value is indeterminate (# is the predicate for indeterminacy):

$$\forall \text{ atomic } \varphi \in \mathcal{T}[\#(\varphi) \leftrightarrow \quad \mathrm{Tr}(\varphi) \wedge \quad F(\varphi)]. \tag{4}$$

Less formally, an atomic theoretical sentence is true iff every rational agent’s probability for it tends to 1 as they approach the limit of inquiry, false iff every rational agent’s probability for it tends to 0 as they approach the limit of inquiry, and indeterminate otherwise.

2.5 TRUTH FOR COMPLEX SENTENCES

We now have a definition of partial truth for the atomic fragment of $\mathcal{L}$. The extension of the truth definition from the atomic sentences to the entire language $\mathcal{L}$ can be carried out in more than one way. The reason is that there is more than one attractive evaluation scheme for partial logic.

One popular such scheme is the *strong Kleene scheme* (Kleene, 1952, Sect. 64). Consider the following ordering $\sqsubset$ on the set of truth values 0 (false), 1 (true) and # (indeterminate): $0 \sqsubset \# \sqsubset 1$. Then the compositional truth clauses of the Kleene valuation scheme V_{SK} take the following form:

∞ $$V_{\mathrm{SK}}(\quad \varphi) = \begin{cases} 1 & \text{if } V_{\mathrm{SK}}(\varphi) = 0, \\ 0 & \text{if } V_{\mathrm{SK}}(\varphi) = 1, \\ \# & \text{if } V_{\mathrm{SK}}(\varphi) = \#; \end{cases}$$

∞ $V_{\mathrm{SK}}(\varphi \vee \psi) = \max\{V_{\mathrm{SK}}(\varphi), V_{\mathrm{SK}}(\psi)\}$;

∞ $V_{\mathrm{SK}}(\exists x \varphi(x)) = \sup\{\varphi(d_i) | i \in \mathbb{N}\}$.

The clauses for the valuation scheme V_{SK} thus provide a way of extending the truth definition to the entire language $\mathcal{L}$.

Another possibility of extending the notion of partial truth to complex sentences is provided by the *supervaluation scheme*. The truth definition for atomic (evidence and theoretical) sentences can be taken to assign an extension and an anti-extension to each predicate of $\mathcal{L}$. But, for partial predicates, some objects of the domain will belong neither to the extension nor to the anti-extension of the predicate. Now call a classical interpretation which is obtained by "filling the gaps" of a partial truth assignment for atomic sentences a *completion* of that truth assignment. For each partial predicate and for each object that according to the partial truth assignment belongs to the gap, the completion will add this object either to the extension, or to the anti-extension of the predicate. This concept will yield an alternative notion V_{SV} of truth for complex formulas:

∞ $V_{SV}(\varphi) = 1 \leftrightarrow V_C(\varphi) = 1$ for all completions V_C;
∞ $V_{SV}(\varphi) = 0 \leftrightarrow V_C(\varphi) = 0$ for all completions V_C;
∞ $V_{SV}(\varphi) = \#$ otherwise.

The strong Kleene scheme has the virtue that it is compositional. The truth value of a disjunction, for instance, is determined by the truth values of the disjuncts. But it has the marked disadvantage that it does not guarantee the truth of all tautologies: if φ is a gappy sentence, then $\varphi\vee\ \ \varphi$ will be just as gappy. While this is not *inconsistent* with anything said so far, some might feel that it does not harmonize very well with the fact that rational agents are required to assign probability 1 to all logical truths right from the beginning. One possible response to this would be to invoke non-classical probability functions such as have been worked out by Weatherson (2003) and (independently and differently) Cantwell (2006); assigning probability 1 to classical tautologies is not a requirement for such probability functions. Another possible response would be to restrict explicitly our definition of truth to contingent sentences. Here we will not attempt to decide which of these approaches (if any) it is best to adopt; we merely want to lay out the options.

One advantage of the supervaluation concept of truth is that it makes all tautologies come out true. Thus, it meshes better with the notion of personal probability on which it is based. On the other hand, it must be noted that the supervaluation concept of truth is not compositional.

There is even a third, more straightforward way in which the notion of partial truth could be extended to the entire language $\mathcal{L}$. Instead of systematically extending the notion of partial truth from atomic to complex sentences using the evaluation schemes V_{SK} or V_{SV}, truth could be defined directly for *all* theoretical sentences of $\mathcal{L}$ on the basis of this generalization of (2):

$$\forall \varphi \in \mathcal{T}\Big[\mathrm{Tr}(\varphi) \leftrightarrow \forall i \in I \lim_{t\to\infty} \mathrm{Pr}_i(\varphi|\wedge A_t) = 1\Big]. \tag{5}$$

Attractive though this may appear, (5) also comes at a cost. As is shown in the Appendix, by adopting (5) we run the risk of ending up with a theory of truth that is ω-inconsistent: a language with a probability assignment can be concocted, which results in the limiting probability value of some sentence $\forall xFx$ being 0 even though the limiting probability value of each of its instances equals 1. Notice that this gives no reason whatsoever to believe that truth will actually be ω-consistent. Indeed, for all we know, the set of sentences that are classified as true by (5) will be ω-consistent for almost any language *cum* accompanying probability assignment.

For some, the mere possibility of ω-inconsistency may be enough to keep them from adopting (5). Let us therefore note that, while from the perspective of a correspondence theorist ω-inconsistency may appear to be a fatal defect – given that it would seem difficult to reconcile the fact expressed by $\forall xFx$ with the facts expressed by the instances of $\forall xFx$ – antirealists, who are not so clearly wedded to an ontology of facts, may be less reluctant to accept the possibility of truth being ω-inconsistent. It is not, of course, as though ω-inconsistency would make every sentence of the language come out true. In this context it is further worth mentioning that there are precedents of *realist* ω-inconsistent theories of truth that are taken seriously in the literature. If one is a deflationist about truth and eschews truth makers, then one might accept an ω-inconsistent theory of truth even as a realist.[8]

In this paper, we officially take a Tarskian stance by keeping object-language and metalanguage separate. It may still be worth sketching how our antirealist partial notion of truth could be extended along Kripkean lines to a self-reflexive notion of truth (cf. Kripke, 1975). In outline, the procedure is as follows. First, one expands the language $\mathcal{L}$ to a semantically closed language $\mathcal{L}_{\mathrm{Tr}}$. This language is obtained by adding the truth predicate to $\mathcal{L}$. In stages, the interpretation of the truth predicate will be improved. At stage 0, we leave the truth predicate completely undetermined: we set both its extension and its anti-extension equal to $\emptyset$. Then we consider, given the relevant family of probability functions and clauses (1)–(4), the collection of sentences which are made true by the valuation scheme V_{SK}. This collection is made the extension of the truth predicate at stage 1. Equally, the collection of sentences that is assigned value 0 is made the anti-extension of the truth predicate at stage 1. The rest of the sentences of $\mathcal{L}_{\mathrm{Tr}}$ are still left undetermined. And so we go on into the transfinite, taking unions at limit stages. Since the evaluation scheme V_{SK} is monotonic, this process eventually reaches a fixed point. Note that the extension and the anti-extension of the truth predicate will not overlap at any stage of the process. The partial model that is reached at the fixed point is an attractive model for the language $\mathcal{L}_{\mathrm{Tr}}$. In a similar way, an attractive model for $\mathcal{L}_{\mathrm{Tr}}$ can be built using the supervaluation scheme.

3. Material adequacy and paradox

We have given two ways of defining antirealist truth for a language. Do these truth definitions satisfy the disquotationalist schema? Are they formally adequate?

Consider either of our definitions of truth for $\mathcal{L}$. Suppose that the collection of sentences that are made definitely true according to the given definition are placed in the extension of the truth predicate, and that the sentences that are made definitely false are placed in its anti-extension. Then according to this definition, the Tarski-biconditionals are at least *weakly* satisfied:

For any sentence $\varphi \in \mathcal{L}$: $\mathrm{Tr}(\varphi)$ holds if and only if φ holds.

More carefully stated, $\mathrm{Tr}(\varphi)$ is true iff φ is true, false iff φ is false, and gappy iff φ is gappy. But the material biconditional $\mathrm{Tr}(\varphi) \leftrightarrow \varphi$ is *gappy* if φ is gappy![9] As to the question of formal adequacy, it will be clear that, since the truth predicate is not part of $\mathcal{L}$, the liar paradox cannot arise.

If, as briefly considered above, antirealist truth for $\mathcal{L}$ is extended to a definition for the self-reflexive language $\mathcal{L}_{\mathrm{Tr}}$, we obtain the weak Tarski-biconditionals for the entire language $\mathcal{L}_{\mathrm{Tr}}$:

For any sentence $\varphi \in \mathcal{L}_{\mathrm{Tr}}$: $\mathrm{Tr}(\varphi)$ holds if and only if φ holds.

The self-reflexive version of the truth definition deals with the liar paradox in the Kripkean way. The liar sentence, which says of itself that it is not true, ends up gappy in all fixed points, so it is judged to be truth-valueless. As a solution to the semantic paradoxes, the present truth definition seems just as satisfactory (or unsatisfactory) as Kripke's theory of truth. In particular, just as the strengthened liar paradox continues to mar Kripke's theory, a similar challenge can be mounted here too: if the liar sentence is judged to be gappy, then in particular it fails to be true, but that is exactly what the sentence says of itself, so, it would seem, the sentence is true after all.

4. Fitch's paradox

Say that a sentential operator $\mathcal{O}$ is *factive* whenever $\mathcal{O}\varphi$ entails φ, and that it *distributes over conjunction* whenever $\mathcal{O}(\varphi \wedge \psi)$ entails both $\mathcal{O}\varphi$ and

$\mathcal{O}\psi$. Fitch (1963) has shown, assuming no more than classical logic, that for any sentential operator $\mathcal{O}$ with both of the aforementioned properties, $\forall\varphi(\varphi \to \Diamond\mathcal{O}\varphi)$ entails $\forall\varphi(\varphi \to \mathcal{O}\varphi)$. This has seemed a huge problem for antirealism, for it has been thought that whatever an antirealist theory of truth was exactly going to look like, it would entail that all truths are knowable (by someone at some time), that is,

$$\forall\varphi(\varphi \to \Diamond \mathrm{K}\varphi).^{10} \tag{6}$$

But, assuming that knowledge is both factive and distributes over conjunction, Fitch's result shows that (6) entails the rather incredible-sounding thesis that all truths are known (by someone at some time), that is,

$$\forall\varphi(\varphi \to \mathrm{K}\varphi), \tag{7}$$

a thesis to which few, if any, antirealists would want to commit themselves. That (6) entails (7) is nowadays commonly referred to as "Fitch's paradox."

However, Fitch's paradox is not a problem for the version of antirealism presented here because, for at least two reasons, our theory does not entail (6). First, (2) is compatible with the supposition that it is impossible (for whatever reasons) for any agent to assign probability 1 to any theoretical truth, and it is reasonable to assume (even if not universally assumed) that knowledge requires probability 1. Second, and regardless of whether knowledge requires probability 1, neither (1) nor (2) ensures that if agents assign probability 1 to some true sentence, they will not be in a Gettier situation with respect to that sentence, and thus will not still fail to know it (on any post-Gettier analysis of knowledge).

5. *Intuitive correctness*

In Section 2 we noted that if we adopted a definition of truth for evidence sentences that renders such a sentence true precisely if at some time an agent assigns probability 1 to it, then our theory of truth would very likely be anemic. And it seems that a theory of truth should not militate too much against common sense by making many sentences that intuitively have a truth value (one way or the other) come out as being truth-valueless. That we adopted (1) instead of the aforementioned more straightforward definition is no guarantee that our theory satisfies this condition; it just prevents the theory from failing to satisfy it too obviously. So, *does* our theory satisfy this condition? That is hard to determine, inasmuch as the only information we possess about the degrees-of-belief functions of the members of our community arises from the assumption that these members are rational agents. Since our definition of rationality is a relatively weak one, this will not help us to answer the question of whether for any, or at least for most, atomic theoretical sentences we deem pre-analytically truth-valued, the probabilities all members of the community assign to them in the limit converge to the same extreme value (we cannot even say whether they converge at all). One response to this problem would be to strengthen the definition of rationality. This would dovetail with complaints Bayesians themselves have raised about the standard Bayesian definition of rationality; that a notion of rationality more substantive than the standard one is needed has been argued by reputed Bayesian authors like Ramsey (1926), Maher (1993) and Joyce (2004). Of course, whether our theory satisfies the present adequacy condition given such a strengthened definition of rationality will depend on the precise nature of the strengthening. Unfortunately, the aforementioned authors do not make any concrete proposals for a strengthening of the definition of rationality and we do not have any concrete suggestions to offer here either.[11]

Meanwhile, antirealists might try to argue, by appeal to the so-called convergence theorems that Bayesians have been able to prove, that from a realist perspective it should seem likely that at least *extensionally*, truth as defined in Section 2 does not differ from realist truth at all, and that the whole difference between the two positions might reside in the respective *explanations* of why the truth predicate has the extension it has. The convergence theorems purport to show that, within certain bounds, choices of prior probabilities are immaterial, as in the long run people's probabilities for a given sentence will converge to one and the same value, however much their prior probabilities for the sentence may diverge. The strongest result of this sort known to date is due to Gaifman and Snir (1982). Roughly, it says that probabilities go to truth values in the limit; so if φ is true, then in the limit (conditional on infinitely many true evidence sentences, so to speak) its probability will almost surely – in the technical sense of this expression – be 1, and if it is false, then in the same limit its probability will almost surely be 0. To see how the Gaifman–Snir result might be relevant to the topic of antirealist truth, notice that, although this result assumes a Tarskian notion of truth, if it holds, and supposing the realist is willing to grant that (1) is at least extensionally correct, then from her perspective, our theory as a whole must also declare true all sentences that are realistically true, and false all sentences that are realistically false. For if a sentence is realistically true (respectively, false), then, by the above result, in the limit all will assign probability 1 (respectively, 0) to it, and so, by our definition, it will be antirealistically true (false) as well. This would be so regardless of which of the options considered in Section 2.5 one chooses, given that all atomic sentences will, under the circumstances considered here, have a determinate truth value – and the right one, from a realist perspective! Thus the realist could not possibly think that our theory is anemic.

But, as we said, the above statement of Gaifman and Snir's result is rough; it in effect hides some important presuppositions. Most notably, the result holds only on the assumption that the evidence sentences *separate* the models of that language, meaning that for any two models there is some evidence sentence that is true in the one and false in the other.[12] Many philosophers will find this assumption implausibly strong, if only because it amounts to denying the so-called Empirical Equivalence Thesis (EET) according to which every theoretical hypothesis has at least one empirically equivalent rival.[13] (In brief, theories are said to be empirically equivalent iff they are accorded the same confirmation-theoretic status in the light of any possible evidence we may receive.) On the other hand, while EET has been regarded as more or less incontrovertible for quite some time, this is no longer true. In the past two decades or so, especially scientific realists have been busy mounting arguments against it, or at least showing that the thesis is entirely unsubstantiated.[14] But we will refrain here from speculating about the prospects of arguing along the above lines for the intuitive correctness, at least from a realist perspective, of antirealist truth as previously defined.

Further, we also said that a theory of truth should entail certain intuitive generalizations concerning truth. For instance, given any theory of truth it should hold that for no sentence both it and its negation are true. Equally, it should hold that if a disjunction is true, then so is at least one of the disjuncts. The former poses no difficulty for our theory. Whether the latter poses a problem may depend on which of the options presented in Section 2.5 is taken for extending the partial truth definitions (1)–(4) to the rest of the language. As intimated in that section, the supervaluation scheme leaves open the possibility that a disjunction is true without either disjunct being true, the strong Kleene scheme does not do so. Here, too, a final assessment of the matter will have to await the development of a more complete account of the rationality conditions for degrees-of-belief functions.

Finally, it will not have been missed that our definition of truth for atomic theoretical sentences assumes that the true evidence sentences that come to be accepted by the community of rational agents do so *in a determinate order*. But, one may wonder, if they had been accepted in some different order, might that have led to the assignment of different truth values? And if so, would that not be counterintuitive? To answer the first question: it follows from standard arguments in probability theory that, given the very minimal assumptions about the probability functions representing the agents' degrees of belief we have made, it is possible that different orderings of the evidence sentences lead to different truth values of the atomic theoretical sentences. To answer the second: it is not clear that this kind of order dependence should bother the antirealist in the least. If truth is a matter of the opinion the community of rational inquirers comes to agree upon, and if these inquirers' opinions happen to be sensitive to the order in which the evidence sentences come to be accepted, then of course truth will be sensitive to that order. Order dependence might run counter to realist intuitions, but from an antirealist perspective it can only be a matter of course.

6. *Truth and the epistemic*

Truth as defined in Section 2 is antirealist insofar as it secures a conceptual connection with the epistemic: truth for evidence sentences is defined in terms of what probabilities appropriately situated rational agents assign or would assign to them, and truth for the remaining sentences of the language is defined recursively in terms of agents' probabilities for the atomic theoretical sentences conditional on more and more true evidence sentences. One may still wonder, however, whether this definition serves the purposes that have motivated philosophers to endorse a specifically antirealist conception of truth.

The single most important motivation is of a meaning-theoretic nature and has forcefully been argued for by Dummett.[15] In a nutshell, the idea is that knowledge of sentence meaning must be ultimately manifestable in a speaker's behavior, and that this requires that a speaker be able to assert a sentence when (or if) its truth conditions are recognized to obtain. Thus – it has seemed – no truth can obtain unrecognizably, that is, all truths must be knowable. As intimated earlier, this does not follow from our theory.

It is important to note, however, that this motivation relies on a view of assertion that makes knowledge the norm of assertion: one ought to assert only what one knows. And it is arguable on grounds entirely unrelated to the realism debate that this requirement is too strong, and that assertion is really governed by the norm that one ought to assert only what is justifiedly credible to one.[16] Once this is recognized, it is easy to show that any theory of truth entails that knowledge of sentence meaning is fully manifestable if it entails the following:

For any contingently true sentence it is possible to obtain evidence
strong enough to make the sentence justifiedly credible, (8)

where for present purposes the designated kind of evidence can simply be taken to be evidence in the standard Bayesian sense – meaning that it raises the sentence's probability – which in addition raises the sentence's probability above a certain threshold value close to 1 (if it was not already above that threshold).[17]

Does our theory entail (8)? Given any (in the present context) reasonable interpretation of the word "possible" in (8) – like "logically possible" or "metaphysically possible" – our theory entails (8) at least when this is restricted to evidence sentences: if an evidence sentence is true according to (1), then for any agent there must be a logically/metaphysically possible world in which he or she assigns probability 1 to it, in which case he or she must have received evidence for it. After all, his or her initial probabilities are strictly coherent, and thus in particular his or her initial probability for the given evidence sentence must have been lower than 1. Moreover, the evidence must be of the right kind, given that, whatever exactly the threshold value for justification may be, it is, by stipulation, lower than 1.

But the theory does *not*, without further assumptions, entail that it is possible to obtain the requisite kind of evidence for any contingently true *theoretical* sentence. As is shown in the Appendix, we can have, for some predicate F and all $j \in \mathbb{N}$, that

$$\lim_{t\to\infty} \Pr_i(Fd_j | \wedge A_t) = 1, \tag{9}$$

and yet also have that

$$\lim_{t\to\infty} \Pr_i(\forall x Fx | \wedge A_t) = 0. \tag{10}$$

If we do have this, then, both by the strong Kleene scheme and by the supervaluation scheme, $\forall xFx$ is true. However, there is no guarantee that we will ever get any evidence for that sentence. Rather, there is a guarantee that in the long run we will obtain evidence strong enough to make its *negation* justifiedly credible.[18]

Naturally, it might be that the more substantial constraints on rational degrees-of-belief functions which, as intimated in Section 5, various Bayesian epistemologists are looking for, will rule out as being irrational (in the more substantial sense) the kind of degrees-of-belief functions that lead to the joint holding of (9) and (10). But perhaps the foregoing just indicates that we should prefer (5) for defining truth for complex sentences, for then (8) is obviously met. In the above case, if $\forall xFx$ is true according to (5), we will have $\lim_{t\to\infty}\Pr_i(\forall xFx | \wedge A_t) = 1$ by definition. More generally, we can then be assured that if a theoretical sentence φ is true, then, given that rational agents are supposed to update probabilities by dint of Bayes's rule, the probability an agent assigns to φ will converge to 1 "in the limit." It follows from this that at some point on the way to the limit, as more and more evidence sentences come to be accepted by the community of inquirers, the probability of any true theoretical sentence will come to exceed the sentence's initial probability (given, again, that initial probabilities are strictly coherent). And, again for the reason that the threshold is lower than 1, the probability assigned to the sentence will also at some point come to exceed that threshold (if it did not do so already). Since it is certainly logically/metaphysically possible that an agent comes to learn enough evidence sentences for the foregoing to happen, it is also possible to obtain the requisite kind of evidence for any true theoretical sentence. Of course, adopting (5) comes at a cost, as we saw, but not one that should be regarded as being intolerably high. The cost–benefit ratio looks even better in light of the present considerations.

7. Putnam's antirealism

To end, we would like to compare our antirealist theory of truth with Putnam's more informal but still somewhat similar view on truth and point to two problems for the latter that the former avoids. Putnam's theory (as we call it for now, despite its professedly informal character) is not in terms of probabilities, but if we equate, for any φ, belief (simpliciter) in φ with assigning probability 1 to φ, then (1) is indeed a restriction to evidence sentences of that theory, which Wright (2000, p. 338) usefully summarizes as: "*P* is true if and only if were *P* appraised under topic-specifically sufficiently good conditions, *P* would be believed."

We start by discussing a problem Plantinga (1982) presented for what he *thought* was Putnam's theory of truth. In Plantinga's interpretation, this is basically the view represented in the citation from Wright, but with "topic-specifically sufficiently good conditions" replaced by "epistemically ideal conditions." So, if *Q* is the sentence "The epistemically ideal conditions hold," then Plantinga supposed Putnam's theory to be this:

$$\forall\varphi(\mathrm{Tr}(\varphi)\leftrightarrow(Q\,\Box\!\!\rightarrow \mathrm{B}\varphi)), \tag{11}$$

where $\mathrm{B}\varphi$ is to be read as "φ is believed by a rational inquirer" or "φ is rationally acceptable" or "φ is agreed upon by all members of the epistemic community" or some such. While such a reading of Putnam's view on truth may have been invited by his early writings on antirealism (such as, most notably, his 1981), in later publications (e.g. Putnam, 1990, 1994) he made it clear that he did not think there was a single set of epistemically ideal conditions under which all truths could be appraised; conditions that count as sufficiently good for the appraisal of one sentence need not count as sufficiently good for the appraisal of another – which is precisely what the word "topic-specifically" in Wright's formulation of Putnam's theory is meant to convey. Thus, not (11) but (12) formally represents Putnam's view:

$$\forall\varphi(\mathrm{Tr}(\varphi) \leftrightarrow (Q_\varphi \mathrel{\Box\!\!\rightarrow} \mathrm{B}\varphi)), \tag{12}$$

with Q_φ meaning that conditions sufficiently good for the appraisal of φ hold. As Wright [2000] showed, however, it takes but some minor changes to the argument underlying Plantinga's problem to arrive at a problem for (12) as well.

The problem Plantinga discovered is that the advocate of (11) is committed to the claim that the epistemically ideal conditions obtain of necessity, that is, to the truth of $\Box Q$. We shall present the argument in natural deduction form here, which requires, apart from the standard introduction and elimination rules: the obvious introduction and elimination rules for the truth predicate; the necessitation rule, which allows us to conclude $\Box\varphi$ from φ provided there are no uncanceled assumptions; the rule that allows us to conclude $\Diamond\varphi$ from φ; and, finally, the following introduction and elimination rules for the subjunctive conditional, which should be uncontroversial:

$$\dfrac{\varphi \qquad \varphi \mathrel{\Box\!\!\rightarrow} \psi}{\psi}\,{\Box\!\!\rightarrow} E \qquad \dfrac{\Box(\varphi \rightarrow \psi)}{\varphi \mathrel{\Box\!\!\rightarrow} \psi}\,{\Box\!\!\rightarrow} I$$

The argument starts by demonstrating that, given (11) as a theory of truth, the supposition $\mathrm{Tr}(Q) \wedge (Q \wedge \quad \mathrm{B}Q)$ leads to inconsistency:

$$\dfrac{\dfrac{\dfrac{\dfrac{\mathrm{Tr}(Q) \wedge (Q \wedge \neg\mathrm{B}Q)}{\mathrm{Tr}(Q)}{\scriptstyle \wedge E} \quad \dfrac{\forall\varphi(\mathrm{Tr}(\varphi) \leftrightarrow (Q \mathrel{\Box\!\!\rightarrow} \mathrm{B}\varphi))}{\mathrm{Tr}(Q) \leftrightarrow (Q \mathrel{\Box\!\!\rightarrow} \mathrm{B}Q)}{\scriptstyle \forall E}}{Q \mathrel{\Box\!\!\rightarrow} \mathrm{B}Q}{\scriptstyle \rightarrow E} \quad \dfrac{\dfrac{\mathrm{Tr}(Q) \wedge (Q \wedge \neg\mathrm{B}Q)}{Q \wedge \neg\mathrm{B}Q}{\scriptstyle \wedge E}}{Q}{\scriptstyle \wedge E}}{\mathrm{B}Q}{\scriptstyle \Box\!\!\rightarrow E} \quad \dfrac{\dfrac{\mathrm{Tr}(Q) \wedge (Q \wedge \neg\mathrm{B}Q)}{Q \wedge \neg\mathrm{B}Q}{\scriptstyle \wedge E}}{\neg\mathrm{B}Q}{\scriptstyle \wedge E}}{\bot}{\scriptstyle \rightarrow E}$$

Call this derivation Π, and note that since, supposedly, (11) holds of conceptual necessity, so that we may put a necessity operator in front of it, we can make use of it also in a necessitated subproof. To arrive at the promised conclusion, $\Box Q$, we then proceed as follows (the unlabelled vertical dots abbreviate some elementary steps, to avoid cluttering of the proof):

$$\dfrac{\dfrac{\dfrac{\mathrm{Tr}(Q) \leftrightarrow (Q \mathrel{\Box\!\!\rightarrow} \mathrm{B}Q) \qquad \dfrac{\dfrac{\dfrac{\dfrac{{}^1[\Diamond(Q \wedge \neg\mathrm{B}Q)] \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{{}^2[Q \wedge \neg\mathrm{B}Q]}{Q}{\scriptstyle \wedge E}}{\mathrm{Tr}(Q)}{\scriptstyle \mathrm{Tr}\, I} \quad {}^2[Q \wedge \neg\mathrm{B}Q]}{\mathrm{Tr}(Q) \wedge (Q \wedge \neg\mathrm{B}Q)}{\scriptstyle \wedge I}}{\Diamond(\mathrm{Tr}(Q) \wedge (Q \wedge \neg\mathrm{B}Q))}{\scriptstyle \Diamond I} \qquad \begin{matrix}[\mathrm{Tr}(Q) \wedge (Q \wedge \neg\mathrm{B}Q)]^3 \\ \vdots\ \Pi \\ \bot\end{matrix}}{\bot}{\scriptstyle 3\Diamond E}}{\bot}{\scriptstyle 2\Diamond E}}{\neg\Diamond(Q \wedge \neg\mathrm{B}Q)}{\scriptstyle 1\rightarrow I}}{\begin{matrix}\vdots \\ \Box(Q \rightarrow \mathrm{B}Q)\end{matrix}}}{Q \mathrel{\Box\!\!\rightarrow} \mathrm{B}Q}{\scriptstyle \Box\!\!\rightarrow I}}{\mathrm{Tr}(Q)}{\scriptstyle \rightarrow E}}{Q}{\scriptstyle \mathrm{Tr}\, E}}{\Box Q}{\scriptstyle \Box I}$$

(As Wright (2000, p. 342n) notes, the application of the necessitation rule in the last step seems superfluous, as it should appear already worrisome enough that the epistemically ideal conditions actually hold.)

Of course this is a problem for (11), a theory of truth that Putnam does *not* endorse. What Wright points out, however, is that if for some sentence P it should be the case that the conditions good enough for its appraisal are identical to those good enough for the appraisal of Q_P, that is, the sentence saying that the conditions for the appraisal of P are good enough, so that Q_P is true if and only if Q_{QP} is, then we would have

$$\mathrm{Tr}(Q_P) \leftrightarrow (Q_{Q_P} \,\Box\!\!\rightarrow \mathrm{B}Q_P) \equiv \mathrm{Tr}(Q_P) \leftrightarrow (Q_P \,\Box\!\!\rightarrow \mathrm{B}Q_P). \tag{13}$$

And that *would* be a problem for (12), because making the substitutions licensed by (13) in the proofs above, and substituting Q_P for Q throughout therein, would yield a proof for the conclusion that the sufficiently good conditions for the appraisal of P obtain of necessity. Although Wright is, as he admits, unable to show that there exists any P for which $Q_P \equiv Q_{QP}$, he rightly remarks that the burden is on Putnam to show that such sentences do not exist – and that may be hard to accomplish. Wright could have added that, even if such sentences do exist, that *need* not be problematic; perhaps there are sentences P for which it is not so hard to accept that sufficiently good conditions for their appraisal necessarily obtain. Here too, however, it would be incumbent on Putnam to show that the foregoing is unproblematic for any sentence of the designated kind (should some exist), which again would seem no easy matter.

Does our version of antirealism escape this problem? It does indeed. For while (1) *almost* has the form of (12), it is restricted to elements of $\mathcal{E}$.[19] And the antirealist should have no difficulty drawing an independently plausible distinction between evidence sentences and the rest of the language which excludes sentences of the form "The circumstances are sufficiently good for the appraisal of φ" from the former class. Arguably, judging whether the circumstances are sufficiently good for the appraisal of this or that sentence will involve judging that one's senses and, at the very minimum, one's mind are working properly; and that is a judgment that would seem to *require* evidence about one's eyesight, one's hearing, the functioning of one's mind, and more perhaps. It certainly is not a sentence attributing an observable property or relationship to observable objects, which we earlier proposed as a reasonable characterization of evidence sentences. We may thus assume that, on our theory, for no sentence φ is $\mathrm{Tr}(Q_\varphi) \leftrightarrow (Q_{Q_\varphi} \,\Box\!\!\rightarrow \mathrm{B}Q_\varphi)$ (or $\mathrm{Tr}(Q_\varphi) \leftrightarrow (Q_\varphi \,\Box\!\!\rightarrow \mathrm{B}Q_\varphi)$) a valid instantiation of (1).[20] As a result, the Plantinga–Wright argument does not apply to (1).

The first problem had to do with the fact that (11) pertains to too many sentences. The second one, now to be discussed, rather has to do with the fact that it seems to pertain to too few sentences. Earlier we considered Putnam's description of the sufficiently good conditions for the appraisal of "There is a chair in my study," which we found to make good sense. But now consider, for instance, the sentence "All ravens are black," and suppose it is true. Then, if (11) is our whole theory of truth, there must be sufficiently good conditions such that, were the sentence to be appraised under those conditions, it would be believed. We find it hard to imagine what those conditions could be: seeing all ravens – past, present and future – in one swoop, and in addition being told (by an oracle, we assume) that these are in fact all ravens, past, present and future? Things would seem even more complicated for "Electrons have negative charge" or "Creutzfeldt–Jakob disease is caused by prions." Moreover, while it is already hard to imagine what sufficiently good conditions for the appraisal of any one of the foregoing sentences could amount to, it is even harder to imagine that such conditions could be generally characterized.[21,22]

One possible response for Putnam would be to make strong idealizations about the community of inquirers, endowing its members with capacities that by far transcend ours. Perhaps it *is* imaginable how for such idealized creatures there can be sufficiently good conditions for the appraisal of any of the aforementioned sentences. As intimated earlier, however, to make this move would be to abandon the arguably most central antirealist tenet, namely, that truth is linked to *our* cognitive capacities.

Another response would be to claim that "Electrons have negative charge" and similar sentences fail to have a truth value. But thereby we would fall short – by a stretch – of satisfying the desideratum that at least many of the sentences we pretheoretically think are truth-valued should come out as indeed being truth-valued on an antirealist (or any other) theory of truth.

Needless to say, this second problem does not arise for our theory either, as the sentences problematic for Putnam are outside the scope of (1). On our theory, the sentence "Electrons have negative charge," being a complex theoretical sentence, can be true without there being sufficiently good conditions for its appraisal.

8. Concluding remarks

Antirealism has so far been a relatively unpopular position. One of the main reasons for this is that it seemed to be beset by a series of quasi-logical difficulties such as Fitch's paradox and Plantinga's argument. Because antirealist theories of truth were for the most part not stated with due precision, it was difficult to gauge accurately the scope of the logical counterarguments. As a consequence, the impression took hold that antirealist truth in general is incoherent. We have been concerned with developing a Peircean conception of truth. While Peirce's antirealist credo, applied to truth, only carries us so far, we hope to have shown that it can be cashed out in a natural and precise way in terms of the key concepts of Bayesian epistemology. If nothing else, the resulting theory (or rather theories, considering the options we left open) has taught us the lesson that we must differentiate between the various quasi-logical difficulties marring antirealist conceptions of truth, and that not every antirealist theory of truth is equally vulnerable to all such objections that have been articulated in the literature.

Igor Douven
Institute of Philosophy, University of Leuven

Leon Horsten
Department of Philosophy, University of Bristol

Jan-Willem Romeijn
Faculty of Philosophy, University of Groningen

ACKNOWLEDGEMENTS

We are greatly indebted to Timothy Williamson for extensive and valuable comments on two previous versions of this paper and for giving us the argument presented in the Appendix. Thanks also to Martin Fischer, David Teira, Christopher von Bülow, Jon Williamson and an anonymous referee for very helpful comments.

APPENDIX

This appendix shows that defining antirealist truth values both for atomic sentences and for logically complex theoretical sentences by means of limiting probability assignments, in the way of (5), may lead to ω-inconsistency. It shows this by providing an example of a probability distribution that, given the said definition, would lead to an ω-inconsistent antirealist truth valuation.[23]

Let the language $\mathcal{L}$ consist of the logical constants , $\wedge$, countably many constants a_t with $t \in \mathbb{N}$, two monadic predicates F and G, both of which can occur in evidence sentences, and the universal quantifier. Next let $Z_{2t} := Ga_{2t}$ and $Z_{2t+1} := \quad Ga_{2t+1}$ for $t \in \mathbb{N}$, and let $A_0 := \top$ and $A_{t+1} := A_t \wedge Fa_t \wedge Z_t$ for $t \in \mathbb{N}$.

Now let Pr* be an arbitrary strictly coherent probability distribution over $\mathcal{L}$. Then we can construct a probability distribution Pr over $\mathcal{L}$ thus:

$$\Pr(A_{t+1}|A_t) = \frac{1}{t+3},$$

$$\Pr(Fa_t \wedge \quad Z_t|A_t) = \frac{1}{t+3},$$

$$\Pr(\quad Fa_t|A_t) = \frac{t+1}{t+3},$$

$$\Pr(X|\quad Fa_t \wedge A_t) = \Pr{}^*(X|\quad Fa_t \wedge A_t), \tag{14}$$

$$\Pr(X|Fa_t \wedge \quad Z_t \wedge A_t) = \Pr{}^*(X|Fa_t \wedge \quad Z_t \wedge A_t), \tag{15}$$

where X is any sentence in $\mathcal{L}$.

By construction, it holds for all $t,\ t' \in \mathbb{N}$ with $t > t'$ that

$$\Pr(Fa_{t'}|A_t) = 1,$$

so that for all t' we have $\lim_{t\to\infty} \Pr(Fa_{t'} \mid A_t) = 1$. But at the same time we have

$$\lim_{t\to\infty} \Pr(\forall x Fx|A_t) \leqslant \lim_{t\to\infty} \Pr(Fa_t|A_t) = \lim_{t\to\infty} \frac{2}{t+3} = 0.$$

If (2) were to hold unrestrictedly for theoretical sentences, then this would be a case of ω-inconsistency – provided, of course, that the probability distribution Pr is acceptable as a basis for determining antirealist truth values, specifically, that it is strictly coherent over $\mathcal{L}$.

To see that it is acceptable as such indeed, first suppose that some logically consistent sentence $X \in \mathcal{L}$ entails A_t for all t. Because X is a finite expression, for some t, X does not contain $a_{t'}$ for any $t' > t$. However, for some $t' > t$, $\mathrm{A}_{t'+1}$ entails $Ga_{t'} \wedge \quad Ga_{t'+1}$. Hence, by assumption, X entails $Ga_{t'} \wedge \quad Ga_{t'+1}$. Because X contains neither $a_{t'}$ nor $a_{t'+1}$, we can substitute $a_{t'+1}$ for $a_{t'}$ in the proof of $Ga_{t'} \wedge \quad Ga_{t'+1}$ from X, thereby deriving $Ga_{t'+1} \wedge \quad Ga_{t'+1}$ from X. But this contradicts our assumption that X is consistent. Therefore, if X is consistent, X cannot entail A_t for all t.

So suppose that X is consistent. Then, because X entails $A_0\ (= \top)$, by the foregoing there must be some t such that X entails A_t but not A_{t+1}. This means that, for some t, X does not entail Fa_t or X does not entail Z_t. If the former, then the sentence $X \wedge A_t \wedge \quad Fa_t$ is consistent, so that $\Pr^*(X \wedge A_t \wedge \quad Fa_t) > 0$, and hence $\Pr^*(X \mid A_t \wedge \quad Fa_t) > 0$, and so, by Equation (14), $\Pr(X \mid A_t \wedge \quad Fa_t) > 0$, so $\Pr(X \wedge A_t \wedge \quad Fa_t) > 0$, so $\Pr(X) > 0$. If the latter, that is, if X does not entail Z_t, the sentence $X \wedge A_t \wedge Fa_t \wedge \quad Z_t$ is consistent, so that $\Pr^*(X \wedge A_t \wedge Fa_t \wedge \quad Z_t) > 0$. One derives that $\Pr(X) > 0$ analogously to how it was just derived, but with $Fa_t \wedge \quad Z_t$ in the place of $\quad Fa_t$ and using Equation (15) instead of Equation (14). Thus, $\Pr(X) > 0$ for every consistent sentence $X \in \mathcal{L}$. In other words, Pr is a strictly coherent probability distribution over $\mathcal{L}$.

NOTES

[1] See Gaifman and Snir (1982, p. 501) for more on axiom (A4), which is a version of countable additivity.

[2] Note that this means that all empirical (i.e. non-logical) sentences receive positive probability. This is possible because probabilities are taken to be defined on sentences, of which there are only denumerably many.

[3] See, for instance, Kemeny (1955), Jeffreys (1961) and Stalnaker (1970).

[4] See in the same vein Lewis (1980, p. 88). But see also Williamson (2007) and Weintraub (2008) for critical discussion.

[5] The operationalist definition of subjective probability originates with Ramsey (1926) and de Finetti (1937); see Gillies (2000) for a very accessible exposition of their views, and for an argument to the effect that operationalism is still the correct view of measurement (or, if you like, of definition) for the social sciences.

[6] It goes without saying that the canonical form of an evidence sentence should not contain any indexicals, or else the sentence might not have a fixed meaning for all circumstances of appraisal. For instance, "My computer makes a humming sound" would mean something different for different agents. Clearly, given our assumption that every object in the domain of discourse is named, we can dispense with indexicals (e.g. we can refer to the computer simply by its name).

[7] But then what about sentences such as "Human kind is being wiped out by a nuclear explosion"? Notice that this should be expected to qualify as an evidence sentence on no plausible definition of evidence sentences.

[8] See Halbach and Horsten (2005).

[9] Of course, it is only from the point of view of the antirealist theory of truth that the Tarski-biconditionals are weakly satisfied. A proponent of a theory of truth according to which there are no truth value gaps, for instance, may be expected to claim that the antirealist theory of truth does not assign the correct extension to the truth predicate. Thus, the mere fact that from the point of view of the antirealist theory the Tarski-biconditionals are weakly satisfied will do nothing to sway the defender of bivalent truth.

[10] The operator K is to be interpreted as "it is known by someone at some time."

[11] Arguably, further rationality constraints on the initial probability assignment Pr are provided by proponents of objective Bayesianism, such as, most notably, Carnap (1950). While for him the further constraints derive, ultimately, from the logical relations between the various sentences of the language, other objective Bayesians, like Jeffreys (1961), Paris (1994) and Jaynes (2003), invoke some version of the Principle of Indifference, or Principle of Minimal Information, typically implemented by means of maximum entropy, to restrict the set of probability assignments that may represent rational degrees of belief. Here we will not comment on the prospects of this program nor on how well its assumptions mesh with the tenets of our antirealist proposal.

[12] Actually the assumption is a bit weaker, namely, that the evidence sentences are "almost everywhere separating," meaning that they separate the models in a class of models of measure 1; see Gaifman and Snir (1982, p. 510) for the details.

[13] See Earman (1992, p. 149 ff).

[14] See, for instance, Leplin (1997) and Kitcher (2001).

[15] See, for instance, Dummett (1976).

[16] See Douven (2006). For defenses of the view that assertion requires knowledge, see Williamson (2000), Adler (2002) and Sundholm (2004), among others.

[17] See Douven (2007, Sect. 5) for the arguments.

[18] Note that, by itself, there is nothing unsettling about this. Epistemologists generally agree that the right account of justification must be of a fallibilist variety, meaning that

it must allow for the possibility that we are sometimes justified in believing something false. Further note that, supposing that the agents will eventually be justified in believing Fd_j for all $j \in \mathbb{N}$, and that justified credibility is closed under logical consequence, there still is no guarantee that any agent will ever be justified in believing $\forall xFx$, as our logic does not contain the ω-rule. But those familiar with discussions pertaining to Kyburg's (1961) Lottery Paradox will know that our assumption that evidence raising a sentence's probability above a given threshold value is sufficient for justified belief in that sentence can only be a simplification (at least if we want to hold on to certain plausible closure conditions on justified belief, which we do). Given a more refined account of justification, it may well be that no element of the set $\{Fd_j \mid j \in \mathbb{N}\}$ will be justifiedly credible to any agent who has a degrees-of-belief function of the kind that leads to (9) and (10), since, relative to such a degrees-of-belief function, the designated set generates an infinite version of the Lottery Paradox, which is no less problematic than the standard finite version (cf. Douven, 2002, Appendix).

[19] Or if we can define generally the sufficiently good conditions for the appraisal of the elements of $\mathcal{E}$, (1) has the even simpler form of (11), again restricted to evidence sentences, of course.

[20] Nor could the sentence "Q will never obtain," which – as Wright (2000, p. 344) points out – Plantinga could also have used to create trouble for the advocate of (11), be validly instantiated in either (11) or (12), once these are restricted to evidence sentences.

[21] And a general characterization is what we need if it is a *definition* of truth that we are after. This may not be Putnam's main concern, who, as intimated at the outset, apparently only had the intention of offering an informal elucidation of truth. But an informal elucidation will do nothing to take away Williamson's also earlier-mentioned complaint that antirealists tend to offer little more than programmatic sketches of their position.

[22] The remarks in this paragraph apply with a vengeance if, like Putnam (1994), one wants to be a direct realist, that is (roughly), maintain that the objects of our experience are not representations of the things surrounding us, but those things themselves. It may be possible to argue that one is directly aware of the chair in one's study, but not – it seems – that one is or could be directly aware of the electrons surrounding one, or of all ravens (past, present and future). See on this also Wright (2000, p. 364).

[23] We owe the example to Timothy Williamson.

REFERENCES

Adler, J. (2002). *Belief's Own Ethics*. Cambridge, MA: MIT Press.

Cantwell, J. (2006). "The Laws of Non-bivalent Probability," *Logic and Logical Philosophy* 15, pp. 163–171.

Carnap, R. (1950). *The Foundations of Probability*. Chicago: University of Chicago Press.

de Finetti, B. (1937). "Foresight: Its Logical Laws, Its Subjective Sources," in H. Kyburg and H. Smokler (eds) *Studies in Subjective Probability*. New York: Krieger, 1980, 2nd edn., pp. 53–118.

Douven, I. (2002). "A New Solution to the Paradoxes of Rational Acceptability," *British Journal for the Philosophy of Science* 53, pp. 391–410.

Douven, I. (2006). "Assertion, Knowledge, and Rational Credibility," *Philosophical Review* 115, pp. 449–485.

Douven, I. (2007). "Fitch's Paradox and Probabilistic Antirealism," *Studia Logica* 86, pp. 149–182.

Dummett, M. A. E. (1976). "What Is a Theory of Meaning? (II)," in G. Evans and J. McDowell (eds) *Truth and Meaning*. Oxford: Clarendon Press, pp. 67–137.

Earman, J. (1992). *Bayes or Bust?* Cambridge, MA: MIT Press.

Fitch, F. B. (1963). "A Logical Analysis of Some Value Concepts," *Journal of Symbolic Logic* 28, pp. 135–142.

Gaifman, H. and Snir, M. (1982). "Probabilities over Rich Languages," *Journal of Symbolic Logic* 47, pp. 495–548.

Gillies, D. (2000). *Philosophical Theories of Probability*. London: Routledge.

Halbach, V. and Horsten, L. (2005). "The Deflationist's Axioms for Truth," in J. C. Beall and B. Armour-Garb (eds) *Deflationism and Paradox*. Oxford: Clarendon Press, pp. 203–217.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.

Jeffreys, H. (1961). *Theory of Probability*, 3rd edn. Oxford: Clarendon Press.

Joyce, J. (2004). "Bayesianism," in A. R. Mele and P. Rawling (eds) *The Oxford Handbook of Rationality*. Oxford: Oxford University Press, pp. 132–155.

Kemeny, J. (1955). "Fair Bets and Inductive Probabilities," *Journal of Symbolic Logic* 20, pp. 263–273.

Kitcher, P. (2001). "Real Realism: The Galilean Strategy," *Philosophical Review* 110, pp. 151–197.

Kleene, S. C. (1952). *Introduction to Metamathematics*. Amsterdam: North-Holland.

Kripke, S. (1975). "Outline of a Theory of Truth," *Journal of Philosophy* 72, pp. 690–716.

Kyburg Jr., H. (1961). *Probability and the Logic of Rational Belief*. Middletown, CT: Wesleyan University Press.

Leplin, J. (1997). *A Novel Defense of Scientific Realism*. Oxford: Oxford University Press.

Lewis, D. (1980). "A Subjectivist's Guide to Objective Chance," in R. Jeffrey (ed.) *Studies in Inductive Logic and Probability*. Berkeley: University of California Press, pp. 263–293. (Reprinted in his Philosophical Papers, Vol. II, Oxford: Oxford University Press, 1986, pp. 83–113; the page reference is to the reprint.)

Maher, P. (1993). *Betting on Theories*. Cambridge: Cambridge University Press.

Paris, J. (1994). *The Uncertain Reasoner's Companion*. Cambridge: Cambridge University Press.

Peirce, C. S. (1978). *Collected Papers*, C. Hartshorne, P. Weiss and A. Burks, eds. Cambridge, MA: Harvard University Press.

Plantinga, A. (1982). "How to Be an Anti-realist," *Proceedings and Addresses of the American Philosophical Association* 56, pp. 47–70.

Putnam, H. (1981). *Reason, Truth and History*. Cambridge: Cambridge University Press.

Putnam, H. (1990). *Realism with a Human Face*. Cambridge, MA: Harvard University Press.

Putnam, H. (1994). "Sense, Nonsense, and the Senses: An Inquiry into the Powers of the Human Mind," *Journal of Philosophy* 91, pp. 445–517.

Ramsey, F. P. (1926). "Truth and Probability," in his *The Foundations of Mathematics*. London: Routledge, 1931, pp. 156–198.

Shope, R. (1978). "The Conditional Fallacy in Contemporary Philosophy," *Journal of Philosophy* 75, pp. 397–413.

Stalnaker, R. (1970). "Probability and Conditionals," *Philosophy of Science* 28, pp. 64–80.

Sundholm, G. (2004). "Antirealism and the Roles of Truth," in I. Niiniluoto, M. Sintonen, and J. Woleński (eds) *Handbook of Epistemology*. Dordrecht: Kluwer, pp. 437–466.

Tarski, A. (1956). "The Concept of Truth in Formalized Languages," in his *Logic, Semantics, Metamathematics*. Oxford: Oxford University Press, pp. 152–278.

Weatherson, B. (2003). "From Classical to Intuitionistic Probability," *Notre Dame Journal of Formal Logic* 44, pp. 111–123.

Weintraub, R. (2008). "How Probable is an Infinite Sequence of Heads? A Reply to Williamson," *Analysis* 68, pp. 247–250.

Williamson, T. (2000). *Knowledge and Its Limits*. Oxford: Oxford University Press.

Williamson, T. (2006). "Must Do Better," in P. Greenough and M. Lynch (eds) *Truth and Realism*. Oxford: Oxford University Press, pp. 177–187.

Williamson, T. (2007). "How Probable is an Infinite Sequence of Heads?" *Analysis* 67, pp. 173–180.

Wright, C. (2000). "Truth as Sort of Epistemic: Putnam's Peregrinations," *Journal of Philosophy* 97, pp. 335–364.