

Antirealist Truth

Igor Douven

Institute of Philosophy, University of Leuven
igor.douven@hiw.kuleuven.be

Leon Horsten

Department of Philosophy, University of Bristol
leon.horsten@bristol.ac.uk

Jan-Willem Romeijn

Faculty of Philosophy, University of Groningen
j.w.romeijn@rug.nl

Abstract

Until now, antirealists have offered sketches of a theory of truth, at best. In this paper, we present an antirealist theory of truth in some formal detail. We show that the theory is able to deal satisfactorily with the problems that are standardly taken to beset antirealism.

According to antirealists, there is an intimate connection between truth and human cognitive capacities which holds by conceptual necessity. While antirealists differ about the exact nature of the connection, no antirealist disputes its conceptual necessity; it distinguishes 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 [1981:56]), 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] seems right to castigate antirealists for, so far at least, failing to offer anything going beyond a merely programmatic sketch of their position. In this paper, we aim to address this lack by taking at least some first steps towards defining a formally precise antirealist theory of truth for a language.

The adequacy conditions for an antirealist theory of truth are partly the same as those for a realist theory of 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 it 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 validate our core intuitions about truth as much as possible. For instance, it should make most, and preferably all, sentences we pretheoretically regard as being truth-valued come out as such. Likewise, it should entail certain

generalizations about truth, such as that a conjunction is true if and only if both of its conjuncts are true. Finally, 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.

In the following, we offer a theory of truth that, as far as we can tell, satisfies the above conditions. We begin, in Section 1, by stating the core of the theory and by addressing some concerns that one might have about it. We then, in Sections 2–5, consider how the theory fares with respect to the above adequacy conditions. Finally, we argue that our theory compares favorably with Putnam’s informal elucidation of antirealist truth, and this not merely on the count of formal precision (Section 6).

1. Antirealist Truth Defined. Our theory can be regarded as being, to some extent, 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]).¹ We 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.

1.1. The language. We are giving a *definition of truth* for a language \mathcal{L} which is supposed to be a regimented language in which empirical scientific theories can be expressed.

\mathcal{L} is a first-order language. It includes the usual logical vocabulary. It also includes mathematical vocabulary, and some non-mathematical vocabulary. 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 in \mathcal{L} : this is considered to be a meta-linguistic notion. And since we plan to reduce truth to degrees of belief or subjective probabilities, the (subjective) probability operator is also considered as a meta-linguistic 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. For convenience, we shall assume a fixed domain $\mathcal{D} = \{d_0, d_1, d_2, \dots\}$ in the following. 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 meta-linguistic sentence variables; we trust that context will suffice to distinguish between the distinct uses.

We further assume that there is a designated part of the language, $\mathcal{E} \subset \mathcal{L}$, such that all and only sentences belonging to \mathcal{E} are apt to report evidence. In Bayesian terms this means that they can receive probability 1 as a direct effect of experience, or at least that rational agents are willing to assign probability 1 to them directly on the basis of their experiences; the other sentences in \mathcal{L} can have their probability altered only mediately, because some evidence sentence receives probability 1. Sentences that are not evidence sentences are called “theoretical sentences”; $\mathcal{T} = \mathcal{L} \setminus \mathcal{E}$ is the class of theoretical sentences. Below we will be more specific about what distinguishes the evidence sentences from the theoretical ones, but for a beginning the above will do.

¹See also Peirce [1978, 5.565]: “Truth is that concordance of an abstract statement with the ideal limit towards which endless investigation would tend to bring scientific belief . . .”

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

1.2. Probability. Roughly corresponding to the Peircean community of “all who investigate,” we assume a community of rational agents. An agent is supposed to have a degrees-of-belief function defined on all sentences of \mathcal{L} , and she is said to be rational iff she satisfies the following three conditions: First, her degrees of belief at all times are representable by a probability function, where a probability function is a function Pr satisfying the following axioms:

- (A1) $0 \leq \text{Pr}(\varphi) \leq 1$ for all $\varphi \in \mathcal{L}$;
- (A2) $\text{Pr}(\varphi) = 1$ for all $\varphi \in \mathcal{L}$ such that φ is a logical truth;
- (A3) $\text{Pr}(\varphi \vee \psi) = \text{Pr}(\varphi) + \text{Pr}(\psi)$ for all $\varphi, \psi \in \mathcal{L}$ such that φ is inconsistent with ψ ;
- (A4) $\text{Pr}(\exists x\varphi(x)) = \lim_{n \rightarrow \infty} \text{Pr}(\bigvee_{i=0}^n \varphi(d_i))$ for all open formulas $\varphi(x) \in \mathcal{L}$ with at most x free.²

Second, her initial probabilities are strictly coherent, that is, before she has obtained any evidence, she assigns probability 1 only to logical truths and thus probability 0 only to logical falsehoods.³ And third, as she receives new evidence, she updates her probabilities by dint of Bayes’s rule. That is, for any given sentence φ , the agent’s new probability for φ after she has become certain of ψ equals 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). As for strict coherence, this has been defended as a *general* requirement of rationality by various authors.⁴ 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?

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 $\varphi \leftrightarrow \text{Tr}(\varphi)$, which Tr satisfies, as will be seen in Section 2. But the use of the truth predicate is clearly dispensable here. For those who have qualms about the

²See Gaifman and Snir [1982:501] for more on axiom (A4), which is a version of countable additivity.

³Note 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.

⁴See, for instance, Kemeny [1955], Jeffreys [1961], and Stalnaker [1970].

⁵The operationalist definition of subjective probability originates with Ramsey [1926] and de Finetti [1937]; see Gillies [2000] for a very accessible exposition of their views (Ch. 4), 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.

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 [1926:161] thought was at least conceivable), then the truth predicate may be equally dispensable for a proper definition of the notion of probability.

1.3. Truth for evidence sentences. The definition of truth is in 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:

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

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} . But, for reasons of simplicity, we stick to (1) in the following.

One possible worry about this notion of truth for evidence sentences 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 [1989], who proposed something very similar to (1) as applying to *all* sentences in the language (more on this in Section 6), 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. vii). 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 a 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 obstructs 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. But this at most suggests the need for a definition with multiple clauses; for instance, one for sentences attributing a *visible* property to an observable entity or process, others for sentences attributing an *audible* 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.⁶

1.4. 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 $\mathcal{E}_{\text{Tr}} \subset \mathcal{E}$ be the set of evidence sentences that are true according to (1), and let $A_t \subseteq \mathcal{E}_{\text{Tr}}$ 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 ; “ $\bigwedge A_t$ ” designates the conjunction of the elements of A_t . Further, let the sequence $\langle A_0, A_1, A_2, \dots \rangle$ 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:

$$(2) \quad \forall \text{ atomic } \varphi \in \mathcal{T} \left[\text{Tr}(\varphi) \leftrightarrow \forall i \in I \lim_{t \rightarrow \infty} \text{Pr}_i(\varphi \mid \bigwedge A_t) = 1 \right],$$

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

$$(3) \quad \forall \text{ atomic } \varphi \in \mathcal{T} \left[\text{F}(\varphi) \leftrightarrow \forall i \in I \lim_{t \rightarrow \infty} \text{Pr}_i(\varphi \mid \bigwedge A_t) = 0 \right].$$

⁶Alternatively, we could define the evidence sentences to be precisely those for which the sufficiently good conditions for their appraisal can be defined. We doubt that by doing so we would stray very far from the class of evidence sentences as circumscribed in terms of observables and observable properties and relations. (And there is certainly no reason to think that if we were to follow the alternative suggestion, all sentences of the language would come to qualify as evidence sentences; see Section 6.)

More than these two clauses is needed, unless we want to preempt the question of whether, for all atomic theoretical sentences in \mathcal{T} , the relevant conditional probability assigned to it by any rational agent i will go either to 1 or to 0 “in the limit.” 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):

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

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.

1.5. 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:

$$\begin{aligned} \cdot V_{SK}(\neg\varphi) &= \begin{cases} 1 & \text{if } V_{SK}(\varphi) = 0; \\ 0 & \text{if } V_{SK}(\varphi) = 1; \\ \# & \text{if } V_{SK}(\varphi) = \#; \end{cases} \\ \cdot V_{SK}(\varphi \vee \psi) &= \max \{V_{SK}(\varphi), V_{SK}(\psi)\}; \\ \cdot V_{SK}(\exists x\varphi(x)) &= \max \{\varphi(d_i) \mid d_i \in \mathcal{D}\}. \end{aligned}$$

The clauses for the valuation scheme V_{SK} 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* (van Fraassen [1966]). 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 object of the domain will neither belong to the extension, nor to the anti-extension of the predicate. Now call a *completion* of a partial truth assignment for atomic sentences a classical interpretation which is obtained by “filling the gaps.” For each partial predicate and for each object which 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:

$$\begin{aligned} \cdot V_{SV}(\varphi) = 1 &\leftrightarrow V_C(\varphi) = 1 \text{ for all completions } V_C; \\ \cdot V_{SV}(\varphi) = 0 &\leftrightarrow V_C(\varphi) = 0 \text{ for all completions } V_C; \\ \cdot V_{SV}(\varphi) &\text{ is undefined otherwise.} \end{aligned}$$

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 $\varphi \vee \neg\varphi$ 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):

$$(5) \quad \forall \varphi \in \mathcal{T} \left[\text{Tr}(\varphi) \leftrightarrow \forall i \in I \lim_{t \rightarrow \infty} \text{Pr}_i(\varphi \mid \bigwedge A_t) = 1 \right].$$

Attractive though this may at first appear, it is shown in the Appendix that we can adopt (5) only on pain of having to accept that the resulting theory of truth might end up being ω -inconsistent: a language with a probability assignment can be concocted which results in the limiting probability value of some sentence $\forall xFx$ to be 0 even though the limiting probability value of each of its instances equals 1. Nothing guarantees that truth will actually *be* ω -inconsistent. 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. Still, the mere possibility of ω -inconsistency may be enough to keep many from adopting (5).⁷

At the same time it must be noticed 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 $\neg\forall xFx$ with the facts expressed by its various instances—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 noting 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; see Halbach and Horsten [2005]. We hasten to add, however, that here we do not wish to defend adopting (5), but merely list it as a possible position for the antirealist that cannot be rejected out of hand.

In this paper, we officially take a Tarskian stance by keeping object-language and meta-language separate. It may still be worth sketching how our antirealist partial no-

⁷See, for instance, McGee [1985] and Yablo [1993], who both argue that ω -inconsistent theories are not really better than inconsistent ones.

tion 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}_{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 is 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}_{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. The partial model that is reached at the fixed point is an attractive model for the language \mathcal{L}_{Tr} . In a similar way, an attractive model for \mathcal{L}_{Tr} can be built using the supervaluation scheme.

2. 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 $\varphi \in \mathcal{L}$: $Tr(\varphi)$ holds if and only if φ holds.

This should be interpreted with care. It means that $Tr(\varphi)$ is true iff φ is true, false iff φ is false, and gappy iff φ is gappy. But the material biconditional $Tr(\varphi) \leftrightarrow \varphi$ is *gappy* if φ is gappy!⁸ 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}_{Tr} , we obtain the weak Tarski-biconditionals for the entire language \mathcal{L}_{Tr} :

For any sentence $\varphi \in \mathcal{L}_{Tr}$: $Tr(\varphi)$ 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 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.

⁸Of 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.

3. 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} which has both of the aforementioned properties, $\forall\varphi(\varphi \rightarrow \diamond\mathcal{O}\varphi)$ entails $\forall\varphi(\varphi \rightarrow \mathcal{O}\varphi)$.⁹ 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,¹⁰

$$(6) \quad \forall\varphi(\varphi \rightarrow \diamond K\varphi).$$

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,

$$(7) \quad \forall\varphi(\varphi \rightarrow K\varphi),$$

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 our theory does not entail (6), for various reasons. For one, an evidence sentence can be true according to (1) even if no agent will ever be in a position good enough to appraise the sentence. For another, (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 generally (even if not universally) accepted that knowledge requires probability 1. For a third, 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). In fact, our position is consistent with the supposition that, whether for these or other reasons, *no* truth is knowable.

4. Intuitive Correctness. In Section 1 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 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

⁹Fitch credited this result to an anonymous referee of an earlier paper which Fitch decided not to publish (Fitch [1963:138 n]). It is now known that the anonymous referee was Alonzo Church; see Salerno [2008].

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

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

Nonetheless, it would be too quick to think that our inability to say anything definitive about our theory in connection with the present adequacy condition suggests the preferability of a realist theory of truth. For, from a realist perspective, it may seem likely that at least *extensionally*, truth, as defined in Section 1 (given either of the options presented in Section 1.5), does not differ from realist truth at all; the whole difference between the two positions might reside in the respective *explanations* of why the truth predicate has the extension it has. Let us make this clear.

Bayesians have been concerned for some time with giving so-called convergence theorems, that is, theorems purporting 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. By far, the strongest result of this sort known to date is due to Gaifman and Snir [1982]. Roughly, Theorem 2.1 of their paper 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 be 1, and if it is false, then in the same limit its probability will be 0. We shall say in a minute why this is rough, but for now notice the *prima facie* relevance of this result to our theory. The Gaifman-Snir result assumes a Tarskian notion of truth to be in place and thus cannot be itself used in a definition of truth. But if the result holds (at least in the foregoing rough form), and if the realist is willing to grant us that (1) is at least extensionally correct in that it assigns the correct truth values to all evidence sentences, then from her perspective our theory as a whole, too, must declare true all sentences that are realistically true and false all sentences that are realistically false. For if a sentence is realistically true (false), then by the above result in the limit all will assign probability 1 (respectively, 0) to it, and so then, by our definition, it will be antirealistically true (false) as well. Note that this will be so regardless of which of the options considered in Section 1.5 is taken, 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.

Our statement of Snir's result was rough, as we said. Most notably, this is so because 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

¹¹Arguably, 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 rationality criteria derive, ultimately, from the logical relations between the various sentences of the language, other objective Bayesians, like Jeffrey [1961], Paris [1994], Jaynes [2003], and Williamson [2007], 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 programme nor on how well its assumptions mesh with the tenets of our antirealist.

is true in the one and false in the other.¹² Until relatively recently, most philosophers would have said that this assumption is implausibly strong, for 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 (Put briefly, 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.) But while EET has been regarded as more or less incontrovertible for quite some time, in the past two decades or so especially scientific realists—whose central commitment is that science aims, and largely succeeds, in uncovering the truth about the world—have been busy mounting arguments against it.¹⁴ The reason for this is quite simply that the thesis has been recognized as one of the chief stumbling blocks for a successful defense of scientific realism.¹⁵ The scientific realists' exact arguments against EET need not detain us here; see, for instance, Douven [2008a] for an overview of the most important ones. What is crucial for our present concerns is that, while, strictly speaking, the *semantic* realist can remain neutral concerning EET, most semantic realists are also *scientific* realists.¹⁶ It thus seems that, from the perspective of most semantic realists, Gaifman and Snir's convergence theorem must give reason to believe that, *extensionally*, realist and antirealist truth may not differ at all.¹⁷

Further, we also said that a theory of truth should entail certain intuitive generalizations concerning truth. For instance, it should hold, given any theory of truth, 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 1.5 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. But of course here too it is worth making the dialectical point that, in view of Gaifman and Snir's result, any realist who doubts EET has reason to think that, extensionally, it will make no difference which option is taken, and that on either theory a disjunction will be true iff at least one of the disjuncts is.

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 come to be accepted *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's degrees of belief we have made, it is possible that different orderings of the

¹²Actually 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 [1982:510] for the details.

¹³See Earman [1992:149 ff]; also Earman [1993] and Douven and Horsten [1998].

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

¹⁵An older trend among scientific realists was to argue that the so-called theoretical virtues, like simplicity, scope, and fecundity, can help us choose between empirically equivalent theories, but so far they have been unable to show that these virtues bear the relation to truth that would seem required to make the said argumentative strategy successful.

¹⁶In fact, the only semantic realist we know of who is not also a scientific realist is van Fraassen; see, for instance, his [1980] for an exposition of his view.

¹⁷For a more extensive discussion of how EET relates to the semantic realism debate, see Douven [2007, Sect. 4].

evidence sentences lead to different truth values of the atomic theoretical sentences.¹⁸ 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. An altogether different response to this worry can be derived again from Gaifman and Snir's paper, this time from their Theorem 2.2. It basically says that, under the same conditions under which the earlier-cited theorem holds, different orderings of the evidence sentences will not affect the assignment of probabilities in the limit—at least this holds with probability 1, meaning that it is logically possible that different orderings *will* affect the assignment of probabilities in the limit, but that the probability that this possibility will materialize is zero. It hence seems that, from the perspective of the realist, the possibility that antirealist truth, as defined according to our proposal, is order-dependent is not one to be taken seriously (and from the perspective of the antirealist it should, for the above-mentioned reason, appear to be little more than a matter of course that this order-dependence may occur, supposing the rational inquirers' degrees-of-belief functions are order-sensitive in the relevant sense).

5. Truth and the Epistemic. Truth as defined in Section 1 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 people's 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.¹⁹ 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 justifiably credible to one.²⁰ 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:

- (8) for any contingently true sentence it is possible to obtain evidence strong enough to make the sentence justifiably credible,

¹⁸This is possible, but certainly not necessary. For instance, consider the sequence $\langle A_0, A_1, A_2, \dots \rangle$ we defined in Section 1.4 and let $[A_t] = A_t \setminus A_{t-1}$ for all $t > 0$. Further, let F_n be the class of functions $f: \langle A_0, A_1, A_2, \dots, A_n, A_{n+1}, \dots \rangle \mapsto \langle A_0, A'_1, A'_2, \dots, A'_n, A_{n+1}, \dots \rangle$ that map $[A_i]$ onto $[A_j]$ for all $i, j \leq n$. Then it is easy to show that, given definitions (2)-(4), truth values of atomic theoretical sentences are invariant under application of any element of $\bigcup_{n \in \mathbb{N}} F_n$ to $\langle A_0, A_1, A_2, \dots \rangle$.

¹⁹See, for instance, Dummett [1976].

²⁰See Douven [2006], [2008b]. The view that assertion requires knowledge has been defended by, among others, Williamson [2000], Adler [2002], DeRose [2002], and Sundholm [2004].

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); see Douven [2007] for the arguments.

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 she assigns probability 1 to it, in which case she must have received evidence for it. After all, her initial probabilities are strictly coherent, and thus in particular 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 1.²¹

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 all $d_j \in \mathcal{D}$ that

$$(9) \quad \lim_{t \rightarrow \infty} \Pr_t(Fd_j \mid \bigwedge A_t) = 1$$

and yet also have that

$$(10) \quad \lim_{t \rightarrow \infty} \Pr_t(\forall xFx \mid \bigwedge A_t) = 0.$$

So, if F is a theoretical predicate, then, both by the strong Kleene and by the supervaluation scheme, $\forall xFx$ is true. However, there is no guarantee that we will ever get any evidence for it. Rather, there is a guarantee that in the long run we will obtain evidence strong enough to make its *negation* justifiedly credible.²²

Naturally, it might be that the more substantial constraints on rational degrees-of-belief functions that, as intimated in Section 4, 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 a dialectically safer response to the above problem, at least for the time being, appeals once more to the Gaifman–Snir result. Although it is logically possible that both (9) and (10)

²¹It may sound strange to say that one can receive evidence for an evidence sentence. But of course if one has come to assign probability 1 to an evidence sentence φ , then one *has* received evidence for φ in the formal sense that there is a sentence ψ —namely, φ itself—to which one has come to assign probability 1 and $\Pr(\varphi \mid \psi) > \Pr(\varphi)$ on one’s initial degrees-of-belief function \Pr .

²²Note 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 justified in believing something false. Further note that, supposing that the agents will eventually be justified in believing Fa_i for all $a_i \in \mathcal{D}$, 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 $\{Fd_j \mid d_j \in \mathcal{D}\}$ 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]).

hold for some theoretical predicate F , if the said result holds, then at least from a realist perspective the antirealist need not fear that this possibility might materialize for any such predicate. In that case, if $\forall xFx$ is realistically true, we will have both (9) and $\lim_{t \rightarrow \infty} \Pr_t(\forall xFx \mid \bigwedge A_t) = 1$. More generally, we can then be assured that if a theoretical sentence φ (of whichever complexity) 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.

It is also worth noting that if (5) is adopted, then no such defensive moves are called for, for then (8) is straightforwardly met. While we have not been defending this option, we did note that it cannot be rejected out of hand.

6. 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 belief (simpliciter) in φ with assigning probability 1 to φ (for any φ), then (1) is indeed a restriction to evidence sentences of that theory, which Wright [2000:338] usefully summarizes as: " P is true if and only if were P appraised under topic-specifically sufficiently good conditions, P would be believed."²³

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 believed Putnam's theory to be this:

$$(11) \quad \forall \varphi (\text{Tr}(\varphi) \leftrightarrow (Q \Box \rightarrow B \varphi)),$$

where $B\varphi$ 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

²³According to Wright [2000:351 f], the variable P should be taken to range over propositions, not sentences, else it would be questionable whether "sufficiently good conditions" can be specified in a non-circular way. For instance—says Wright—it would certainly be part of the sufficiently good conditions for the appraisal of "Somebody is standing behind you" to turn around and look. But, having turned around, the sentence would need re-expression. Yet it would be absurd to say that "Somebody is standing behind you" is not verifiable for that reason. This is unconvincing, however, as it would seem reasonable to suppose that the truth of a sentence is to be evaluated in a context (see, e.g., Visser [1989:627]). And we are perfectly able to verify that the sentence to be evaluated in Wright's example is true in context c , say, even if this requires us to be in a context different from c .

(As Wright [2000:342 n] 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 hold actually.)

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_{Q_P} is, then we would have

$$(13) \quad \text{Tr}(Q_P) \leftrightarrow (Q_{Q_P} \Box \rightarrow B Q_P) \equiv \text{Tr}(Q_P) \leftrightarrow (Q_P \Box \rightarrow B Q_P).$$

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 $Q_P \equiv Q_{Q_P}$, 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 they 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} .²⁴ 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 $\text{Tr}(Q_\varphi) \leftrightarrow (Q_{Q_\varphi} \Box \rightarrow B Q_\varphi)$ (or $\text{Tr}(Q_\varphi) \leftrightarrow (Q_\varphi \Box \rightarrow B Q_\varphi)$) a valid instantiation of (1).²⁵ 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

²⁴Or 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.

²⁵Nor could the sentence “ Q will never obtain,” which—as Wright [2000:344] 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.

being told (by an oracle, we assume) that these are in fact all ravens, past, present, and future? Things would even seem more complicated for “Electrons have negative charge” or “Creutzfeldt-Jakob disease is caused by prions.” Moreover, if 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.^{26,27}

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. (For instance, we think it is imaginable what sufficiently good conditions for the appraisal of “All ravens are black” are for the Tralfamadorians in Kurt Vonnegut’s novel *Slaughterhouse-Five*, who can look at all moments in time, past, present, and future, the way we can look at a landscape or a mountain.) 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 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.

7. 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 articulated 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 quasi-logical dif-

²⁶And 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.

²⁷The 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). Wright [2000:364] briefly hints at the tension between Putnam’s view on truth (or actually on what Wright thinks of as an improvement on Putnam’s view on truth) and direct realism. Our theory of truth, which defines truth differently for different segments of the language, might exactly yield the “mixed” position the necessity of which Wright sees as arising from that tension.

faculties 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.

We reemphasize that the foregoing should be thought of as constituting only the first steps towards a full-fledged antirealist theory of truth. Among the further steps to be taken should certainly be the extension of the theory to more inclusive fragments of our language. Also, presently at some junctures the theory is defended by appeal to a result from Gaifman and Snir that is acceptable from a realist, but not from an antirealist, viewpoint. Thereby the dialectical situation is somewhat reminiscent of the days of the first completeness proofs for intuitionistic logic, which, because they assumed classical logic at the meta-level, were classically, but not intuitionistically, acceptable. We hope that, just as later intuitionistically acceptable completeness proofs were given, future research will lead to a justification of our theory of truth that is fully acceptable to the antirealist as well.

Acknowledgements. Earlier versions of this paper were presented at the ESF Exploratory Workshop “Applied Logic in the Methodology of Science” held at the University of Bristol, at the First Meeting in Philosophy, Probability and Scientific Method held at the University of Valencia, at the LERU meeting “Rationality and the Choice of Logic” at the University of Leyden, and at a one-day conference on truth held at Erasmus University Rotterdam. We are grateful to the organizers of these events—respectively Alexander Bird and Hannes Leitgeb, Valeriano Iranzo, Göran Sundholm, and Fred Muller—for the invitations and to the various audiences for stimulating questions and remarks. We are also grateful to David Teira, Christopher von Bülow, and Jon Williamson for very helpful comments on a draft of this paper. Our greatest debt is to Timothy Williamson, both for extensive and valuable comments on a previous version and for giving us the argument presented in the Appendix.

Appendix

This appendix shows that defining antirealist truth values for both atomic sentences and logically complex sentences by means of limiting probability assignments may lead to ω -inconsistency. It does so by providing an example of an antirealist truth valuation that is ω -inconsistent.²⁸

Let the language \mathcal{L} consist of the logical constants \neg, \wedge , countably many constants a_t with $t > 0$, two monadic predicates F and G , and the universal quantifier $\forall t$. For convenience we will abbreviate $E_t^1 = Ga_t$ and $E_t^0 = \neg Ga_t$. Let $s \in 2^\Omega$ be an infinite binary sequence, with $s(t)$ the t -th element of this sequence. Then define

$$S = \{s \mid \forall t > 0 \exists t' > t: s(t) \neq s(t')\}.$$

In words, S is the set of all sequences that do not end in an unbroken infinite string of either 0's or 1's. Next define the following set of sequences of evidence sentences:

$$\begin{aligned} A_1 &= \top, \\ \bigwedge_{t+1} A_{t+1}^s &= \bigwedge_t A_t \wedge Fa_t \wedge E_t^{s(t)}, \\ \mathcal{A} &= \{\bigwedge_t A_t^s \mid s \in S\}. \end{aligned}$$

²⁸We owe the example to Timothy Williamson.

So the evidence sentences $Fa_t \wedge E_t^{s(t)}$ are entailed by $\bigwedge_{t+1} A_{t+1}^s$ but not by $\bigwedge_t A_t^s$. Note further that there are uncountably infinitely many sequences in the set \mathcal{A} , because there are uncountably infinitely many sequences $s \in S$.

Let \Pr_C be a strictly coherent probability distribution over the language \mathcal{L} . For any sequence of evidence sentences $\bigwedge_t A_t^s \in \mathcal{A}$, construct the probability distribution \Pr over the language \mathcal{L} in the following way:

$$\begin{aligned} \Pr(Fa_t \wedge E_t^{s(t)} \mid \bigwedge_t A_t^s) &= \frac{1}{t+2}, \\ \Pr(Fa_t \wedge E_t^{1-s(t)} \mid \bigwedge_t A_t^s) &= \frac{1}{t+2}, \\ \Pr(\neg Fa_t \mid \bigwedge_t A_t^s) &= \frac{t}{t+2}, \\ (14) \quad \Pr(X \mid \bigwedge_t A_t^s \wedge Fa_t \wedge E_t^{1-s(t)}) &= \Pr_C(X \mid \bigwedge_t A_t^s \wedge Fa_t \wedge E_t^{1-s(t)}), \\ (15) \quad \Pr(X \mid \bigwedge_t A_t^s \wedge \neg Fa_t) &= \Pr_C(X \mid \bigwedge_t A_t^s \wedge \neg Fa_t), \end{aligned}$$

where X is any sentence in the language \mathcal{L} . Now for $t > t'$ we have that

$$\Pr(Fa_{t'} \mid \bigwedge_t A_t^s) = 1,$$

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

$$\lim_{t \rightarrow \infty} \Pr(\forall t' Fa_{t'} \mid \bigwedge_t A_t^s) \leq \lim_{t \rightarrow \infty} \Pr(Fa_t \mid \bigwedge_t A_t^s) = \lim_{t \rightarrow \infty} \frac{2}{t+2} = 0.$$

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

To show that it is acceptable indeed, we first prove that an arbitrary consistent sentence X cannot entail $\bigwedge_t A_t^s$ for all s and t . First note that for some specific s we may introduce an atomic sentence $S^s = \lim_{t \rightarrow \infty} \bigwedge_t A_t^s$, so that the sentence $\bigwedge_t A_t^s$ is entailed by S^s for each t . But in the language \mathcal{L} there can be only countably many such atomic sentences. Now choose s^* such that S^{s^*} is not included in \mathcal{L} . We assume that this is the sequence of evidence sentences.

For the sake of argument, suppose that some consistent X entails $\bigwedge_t A_t^{s^*}$ for all t . Then because X is a finite expression in the language, there is an N such that for all $t > N$, the constants a_t do not appear in X . However, since $s^* \in S$ there are $t, t' > N$ such that $s^*(t) = 1$ and $s^*(t') = 0$, so that $Fa_t \wedge Ga_t$ and $Fa_{t'} \wedge \neg Ga_{t'}$. Hence, by assumption, X entails $Ga_t \wedge \neg Ga_{t'}$, but because X does not contain a_t or $a_{t'}$, we can substitute a_t for $a_{t'}$, and derive $Ga_t \wedge \neg Ga_t$ from X . Therefore, if X is consistent, it cannot entail $\bigwedge_t A_t^{s^*}$ for all t .

Because $A_1 = \top$, the sentence X entails A_1 , and so there is some $t > 1$ such that X entails $\bigwedge_t A_t$ but not $\bigwedge_{t+1} A_{t+1}$. (For sake of brevity we are here suppressing the superscript s^* in $\bigwedge_t A_t^{s^*}$.) This is because X does not entail Fa_t , or because X does not entail $E_t^{s^*(t)}$ (or both). In the former case, the sentence $\bigwedge_t A_t \wedge \neg Fa_t \wedge X$ is consistent, so that $\Pr_C(\bigwedge_t A_t \wedge \neg Fa_t \wedge X) > 0$, and hence $\Pr_C(X \mid \bigwedge_t A_t \wedge \neg Fa_t) > 0$, hence $\Pr(X \mid \bigwedge_t A_t \wedge \neg Fa_t) > 0$ because of Equation (15). We have $\Pr(\bigwedge_t A_t \wedge \neg Fa_t) = t/(t+2) \prod_{i=1}^{t-1} 1/(i+2) > 0$, so that $\Pr(\bigwedge_t A_t \wedge \neg Fa_t \wedge X) > 0$ and hence $\Pr(X) > 0$.

In the case that X does not entail $E_t^{s^*(t)}$, the sentence $\bigwedge_t A_t \wedge Fa_t \wedge E_t^{1-s^*(t)} \wedge X$ is consistent, so that $\Pr_C(\bigwedge_t A_t \wedge Fa_t \wedge E_t^{1-s^*(t)} \wedge X) > 0$. We can now derive that $\Pr(X) > 0$

analogously to how it is derived above, with $Fa_t \wedge E_t^{1-s^*(t)}$ in the place of $\neg Fa_t$ and using Equation (14) instead of Equation (15). Thus any consistent X has $\Pr(X) > 0$.

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* 15:163-171.
- 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 ed.), pp. 53-118.
- DeRose, K. [2002] "Assertion, Knowledge, and Context," *Philosophical Review* 111:167-203.
- Douven, I. [2002] "A New Solution to the Paradoxes of Rational Acceptability," *British Journal for the Philosophy of Science* 53:391-410.
- Douven, I. [2006] "Assertion, Knowledge, and Rational Credibility," *Philosophical Review* 115:449-485.
- Douven, I. [2007] "Fitch's Paradox and Probabilistic Antirealism," *Studia Logica* 86:149-182.
- Douven, I. [2008a] "Underdetermination," in S. Psillos and M. Curd (eds.) *The Routledge Companion to Philosophy of Science*, London: Routledge, pp. 292-301.
- Douven, I. [2008b] "Assertion, Moore, and Bayes," *Philosophical Studies*, in press.
- Douven, I. and Horsten, L. [1998] "Earman on Underdetermination and Empirical Indistinguishability," *Erkenntnis* 49:303-320.
- 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.
- Earman, J. [1993] "Underdetermination, Realism, and Reason," in P. French, T. Uehling, Jr., and H. Wettstein (eds.) *Midwest Studies in Philosophy*, Vol. XVIII, Notre Dame: University of Notre Dame Press, pp. 19-38.
- Fitch, F.B. [1963] "A Logical Analysis of Some Value Concepts," *Journal of Symbolic Logic* 28:135-142.
- Gaifman, H. and Snir, M. [1982] "Probabilities over Rich Languages," *Journal of Symbolic Logic* 47:495-548.
- 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 ed.), 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.
- Kaplan, M. [1981] "A Bayesian Theory of Rational Acceptance," *Journal of Philosophy* 78:305-330.
- Kemeny, J. [1955] "Fair Bets and Inductive Probabilities," *Journal of Symbolic Logic* 20:263-273.

- Kitcher, P. [2001] "Real Realism: The Galilean Strategy," *Philosophical Review* 110:151–197.
- Kleene, S. C. [1952] *Introduction to Metamathematics*, Amsterdam: North-Holland.
- Kripke, S. [1975] "Outline of a Theory of Truth," *Journal of Philosophy* 72:690–716.
- Kvanvig, J. [2006] *The Knowability Paradox*, Oxford: Oxford University Press.
- 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.
- Maher, P. [1993] *Betting on Theories*, Cambridge: Cambridge University Press.
- McGee, V. [1985] "How Truthlike Can a Predicate Be?" *Journal of Philosophical Logic* 14:399–410.
- Paris, J. [1994] *The Uncertain Reasoner's Companion*, Cambridge: Cambridge University Press.
- Peirce, C. S. [1978] *Collected Papers* (edited by C. Hartshorne, P. Weiss, and A. Burks), Cambridge MA: Harvard University Press.
- Plantinga, A. [1982] "How to Be an Anti-Realist," *Proceedings and Addresses of the American Philosophical Association* 56:47–70.
- 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* 91:445–517.
- Ramsey, F. P. [1926] "Truth and Probability," in his *The Foundations of Mathematics*, London: Routledge, 1931, pp. 156–198.
- Salerno, J. [2008] "Knowledge Noir: 1945–1963," in J. Salerno (ed.) *New Essays on the Knowability Paradox*, Oxford: Oxford University Press, in press.
- Stalnaker, R. [1970] "Probability and Conditionals," *Philosophy of Science* 28:64–80.
- 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.
- Tennant, N. [1990] *Natural Logic*, Edinburgh: Edinburgh University Press.
- van Dalen, D. [1994] *Logic and Structure* (3rd ed.), New York: Springer.
- van Fraassen, B. C. [1966] "Singular Terms, Truth-value Gaps and Free Logic," *Journal of Philosophy* 63:481–495.
- van Fraassen, B. C. [1980] *The Scientific Image*, Oxford: Clarendon Press.
- Visser, A. [1989] "Semantics and the Liar Paradox," in D. Gabbay and F. Guenther (eds.) *Handbook of Philosophical Logic* (Vol. IV), Dordrecht: Kluwer, pp. 617–706.
- Weatherson, B. [2003] "From Classical to Intuitionistic Probability," *Notre Dame Journal of Formal Logic* 44:111–123.
- Williamson, J. [2007] "Motivating Objective Bayesianism: From Empirical Constraints to Objective Probabilities," in W. L. Harper and G. R. Wheeler (eds.) *Probability and Inference: Essays in Honor of Henry E. Kyburg Jr.*, London: College Publications, pp. 155–183.
- 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.
- Wright, C. [2000] "Truth as Sort of Epistemic: Putnam's Peregrinations," *Journal of Philosophy* 97:335–364.

Yablo, S. [1993] "Paradox Without Self-reference," *Analysis* 53:251-252.