Antirealist Truth

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

Leon Horsten
Institute of Philosophy, University of Leuven
leon.horsten@hiw.kuleuven.be

Jan-Willem Romeijn
Faculty of Philosophy, University of Groningen
j.w.romeijn@rug.nl

Abstract

Antirealists have hitherto offered at best sketches of a theory of truth. This paper presents an antirealist theory of truth in some formal detail. It is shown that the theory is able to deal satisfactorily with some problems that are standardly taken to beset antirealism.

According to antirealists, there is an intimate connection between truth and human cognitive capacities which holds of conceptual necessity. While antirealists differ about the exact nature of the connection, the point of conceptual necessity is undisputed; it distinguishes the antirealist conception of truth from a realist one accompanied by some methodological view to the effect that, by natural selection or just by good fortune perhaps, 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 hope to do better 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 as much as possible validate our core intuitions about truth. 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 disjunction can only be true if at least one of its disjuncts is. 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 that the theory plausibly satisfies the epistemic and meaning-theoretic 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 addressing some worries 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 [1878:155]). 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 $\mathcal{D} = \{d_0, d_1, d_2, \ldots\}$ in the following. When we speak of the sentences of $\mathcal{L}$, we throughout 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.

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) $0 \leqslant \Pr(\varphi) \leqslant 1$ 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 inconsistent with ψ;
(A4) $\Pr(\exists x \varphi(x)) = \lim_{n \to \infty} \Pr(\bigvee_{i=0}^{n} \varphi(d_i))$ for all open formulas $\varphi(x) \in \mathcal{L}$ with at most x free.[1]

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.[2] 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 probability 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.[3] As such it is problematic, however, as strict coherence is incompatible with learning by means of Bayes's rule, given that this rule only 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.[4] Succinctly, one's probability for φ can be interpreted as the maximum price (expressed in Euro-cents) 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, as will be seen shortly. 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

[1] See Gaifman and Snir [1982:501] 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] 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 (Ch. 4), and for an argument to the effect that operationalism is still the correct view of measurement (or, if one likes, if meaning) for the social sciences.

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, then, 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 nonempty set I be the community of rational agents, and let Pr_i^t be person i's probability function at time t. Then a very simple truth definition for $\mathcal{E}$ would be this:

$$\forall \varphi \in \mathcal{E}\ \left[\mathrm{Tr}(\varphi) \ \leftrightarrow\ \exists i \in I\ \exists t\ \mathrm{Pr}_i^t(\varphi) = 1\right]. \tag{1}$$

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 never occur that $\mathrm{Pr}_i^t(\varphi) = \mathrm{Pr}_j^{t'}(\neg\varphi) = 1$, for some $\varphi \in \mathcal{E}$, some agents i and j, and some times t and t'. 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 right-hand side of (1), like "for all agents/most agents/the majority of agents, there is a time at which they will assign probability 1 to φ" 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 seem to share (1)'s defect:

$$\forall \varphi \in \mathcal{E}\ \left[\mathrm{Tr}(\varphi) \ \leftrightarrow\ \left\{\begin{array}{l}\text{for any } i \in I \text{ and any time } t\text{, if } i \text{ were at } t \text{ in circum-}\\ \text{stances sufficiently good for the appraisal of } \varphi\text{, then } i\\ \text{would at } t \text{ assign probability 1 to } \varphi.\end{array}\right\}\right] \tag{2}$$

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 (2) 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 (2) 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 (2) 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 (2) 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 apply already 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.[5]

1.4. Truth for atomic theoretical sentences. There are several ways in which the above truth definition for evidence sentences can be extended to a truth definition for the entire language $\mathcal{L}$.

We first extend the truth definition to atomic theoretical sentences. Let $\mathcal{E}_{\mathrm{Tr}} \subset \mathcal{E}$ be the set of evidence sentences that are true according to (2), and let $K_t \subseteq \mathcal{E}_{\mathrm{Tr}}$ be the set of evidence sentences that are known to the community of rational agents at time t. It is assumed that at any given time there are only finitely many evidence sentences known, so that K_t is finite for all t. Next we define a sequence $\langle K_0, K_1, K_2, \dots \rangle$ of subsets of $\mathcal{E}_{\mathrm{Tr}}$, as follows: (i) $K_0 = \varnothing$; (ii) $K_t \subseteq K_{t+1}$; (iii) $\bigcup_{t \leqslant \omega} K_t = \mathcal{E}_{\mathrm{Tr}}$. Further, $\bigwedge K_t$ is the conjunction of the elements of K_t. Then truth for atomic sentences in $\mathcal{T}$ is defined thus:

$$\text{(3)} \qquad \forall \text{ atomic } \varphi \in \mathcal{T} \ \Big[\mathrm{Tr}(\varphi) \leftrightarrow \forall i \in I \ \lim_{t \to \infty} \mathrm{Pr}_i(\varphi \mid \textstyle\bigwedge K_t) = 1\Big].$$

Unless we want to preempt the question whether, for all atomic sentences in $\mathcal{T}$, the relevant conditional probability assigned to it by any rational agent i will go either to 1

[5] Alternatively, 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.)

or to 0 “in the limit,” more than (3) is needed. For suppose for some such φ convergence in the sense specified here does not occur; perhaps the limit of the relevant conditional probability has some other value than 1 or 0 for some or even all agents. Then from (3) it follows that φ is not true. Does that mean it is false? Not if we want to maintain—as would seem desirable to do—that a sentence is false iff its negation is true. For if φ is not true for the reason just given, then neither can $\neg\varphi$ be true: if it were the case that $\lim_{t\to\infty} \Pr_i(\neg\varphi \mid \bigwedge K_t) = 1$ for all agents i, then, given that they are assumed to be rational, it must also be that $\lim_{t\to\infty} \Pr_i(\varphi \mid \bigwedge K_t) = 0$ for all i, contrary to our supposition. The following two definitions will allow us to distinguish between false sentences and sentences whose truth value is indeterminate (F is the falsity predicate, # the predicate for indeterminacy):

$$(4) \qquad \forall \text{ atomic } \varphi \in \mathcal{T} \ \Big[\mathrm{F}(\varphi) \leftrightarrow \forall i \in I \lim_{t\to\infty} \Pr_i(\varphi \mid \textstyle\bigwedge K_t) = 0\Big];$$

$$(5) \qquad \forall \text{ atomic } \varphi \in \mathcal{T} \ [\#(\varphi) \leftrightarrow \neg\mathrm{Tr}(\varphi) \wedge \neg\mathrm{F}(\varphi)].$$

A comment on this: The above assumes that for every true evidence sentence, there is some time t such that the sentence is known to the community of rational inquirers. This seems to involve quite a lot of idealization: the human race will have a limited time of existence and so it seems reasonable to believe that there exist evidence sentences that are true by (2) but that will never be known by anyone. A first thing to note is that this kind of idealization is not at all foreign to the antirealist conception of truth. It is clearly present, for instance, in Peirce’s earlier-cited idea of an opinion the rational inquirers will ultimately come to agree upon. The “ultimately” clearly suggests an endpoint of all inquiry, but it seems unrealistic to suppose that there will be such an endpoint. The same might be true of the agreement that is supposed to be reached by the inquirers. Putnam [1981], in presenting his antirealist conception of truth, is even more explicit on the point of idealization by calling his view (to be discussed in section 6) an “idealization theory of truth” (p. 56), where the idealization involved is likened to that involved in the assumption of frictionless planes in mechanics (p. 55). So we should expect antirealists to have no difficulty countenancing the above clauses.

Even so, it may be good to point to variants of them that avoid the said idealization. Suppose, plausibly, that inquiry will end some day, and that hence there will be some finite set K_n of true evidence sentences such that the elements of K_n will be all the true evidence sentences the community of rational inquirers will ever come to know. Furthermore, let $\mathbf{t}$ be some value close to 1; one could think here of a threshold for rational acceptance that some authors have proposed.[6] Then consider the following variant of (3):

$$(6) \qquad \forall \text{ atomic } \varphi \in \mathcal{T} \ \Big[\mathrm{Tr}(\varphi) \leftrightarrow \exists t \forall t' \big(t \leqslant t' \leqslant n \to \forall i \in I \Pr_i(\varphi \mid \textstyle\bigwedge K'_t) > \mathbf{t}\big)\Big].$$

Less formally, an atomic theoretical sentence is true precisely if, for some stage of inquiry, all rational inquirers assign the sentence a probability above the threshold conditional on the evidence sentences known at that stage of inquiry, and they also assign it a probability above the threshold conditional on the evidence sentences known at any later stage of inquiry. This variant of (3), with corresponding variants of (4) and (5), may serve the antirealist’s purposes as well as the more idealized clauses. However, whether

[6] See, for instance, Kyburg [1970], Foley [1992], and Achinstein [2001].

this is really so is topic for another occasion; in the following we will stick to clauses (3)–(5).[7]

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 several ways. The reason is that there are various attractive evaluation schemes for partial logic.

One of the most popular evaluation schemes for partial logic is the *Strong Kleene Scheme*. 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:

- $V_{SK}(\neg\varphi) = 0$ if $V_{SK}(\varphi) = 1$;
  $V_{SK}(\neg\varphi) = 1$ if $V_{SK}(\varphi) = 0$;
  $V_{SK}(\neg\varphi) = \#$ if $V_{SK}(\varphi) = \#$;
- $V_{SK}(\varphi \vee \psi) = \max\{V_{SK}(\varphi), V_{SK}(\psi)\}$;
- $V_{SK}(\exists x\varphi(x)) = \max\{\varphi(d_i) \mid d_i \in \mathcal{D}\}$.

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*. 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:

- $V_{SV}(\varphi) = 1 \leftrightarrow V_C(\varphi) = 1$ for all completions V_C;
- $V_{SV}(\varphi) = 0 \leftrightarrow V_C(\varphi) = 0$ for all completions V_C;
- $V_{SV}(\varphi)$ is undefined 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 $\varphi \vee \neg\varphi$ will be just as gappy. And this does not seem to harmonize well with the fact that the notion of truth is supposed to be intimately related to personal probability, for it is a requirement on probability functions that they assign probability 1 to all logical truths. One possible response to this would be to invoke non-classical probability functions such as have been worked out by Weatherson [2003]

[7] Another issue worth investigating is whether the variant-clauses suggested here could serve as a basis for a formalization of Wright's [1992] superassertibility view of truth. At least given a justified credibility account of assertion (see section 5), there seems to be an intuitively close connection between the former and the latter.

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 empirical 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 flip side of this, it must be noted that the supervaluation concept of truth is not compositional.

There is a third, even more straightforward way in which the notion of partial truth can 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}, we can define truth directly for *all* theoretical sentences of $\mathcal{L}$ on the basis of a generalization of (3):

$$\forall \varphi \in \mathcal{T} \left[\mathrm{Tr}(\varphi) \leftrightarrow \forall i \in I \lim_{t \to \infty} \mathrm{Pr}_i(\varphi \mid \bigwedge K_t) = 1 \right]. \tag{7}$$

This will yield a notion of truth which is not necessarily compositional, but does judge all tautologies to be true. Still, it is not necessarily extensionally equivalent with truth as defined via V_{SV}. For instance, it is easy to see that we might have, for some theoretical sentences φ and ψ, (i) $\lim_{t\to\infty} \mathrm{Pr}_i(\varphi \mid \bigwedge K_t) \neq 1$ for some agents i, (ii) $\lim_{t\to\infty} \mathrm{Pr}_i(\psi \mid \bigwedge K_t) \neq 1$ for some agents i, (iii) $\lim_{t\to\infty} \mathrm{Pr}_i(\varphi \vee \psi \mid \bigwedge K_t) = 1$ for all agents i, even though (iv) $\nvDash \varphi \vee \psi$.

At the outset of this paper we have officially taken a Tarskian stance by keeping object-language and meta-language 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 a family of probability functions), 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. Similarly, 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. 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.

2. Material Adequacy and Paradox. We have given three ways of defining antirealist truth for a language. Do these truth definitions satisfy the disquotationalist schema?

Consider one of the three antirealist definitions of truth for $\mathcal{L}$. Suppose that the collection of sentences that are made definitely true 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.

This should be interpreted with care. It means that $T(\varphi)$ is true if φ is true, false if φ is false, and gappy if φ is gappy. But the material biconditional $T(\varphi) \leftrightarrow \varphi$ is gappy if φ is gappy!

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 ends up gappy in all fixed points, so it is judged to be truth-valueless.

So, as a solution to the semantic paradoxes, the present truth definitions are 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.

3. Fitch's Paradox. Say that a sentential operator $\mathcal{O}$ is *factive* whenever $\mathcal{O}\varphi$ entails φ, and say 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 $\forall\varphi(\varphi \rightarrow \mathcal{O}\varphi)$.[8] 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 had to entail that all truths are knowable (by someone at some time), that is,[9]

(8) $$\forall\varphi(\varphi \rightarrow \Diamond\,\mathrm{K}\varphi).$$

But, assuming that knowledge is both factive and distributes over conjunction, Fitch's result shows that (8) entails the rather incredible-sounding thesis that all truths are known (by someone at some time), that is,

(9) $$\forall\varphi(\varphi \rightarrow \mathrm{K}\varphi),$$

a thesis to which few, if any, antirealists would want to commit themselves. That (8) entails (9) is nowadays commonly referred to as "Fitch's Paradox."

Is this a problem for our version of antirealism, too? The answer is that it is not, as our theory does not entail (8). First, most philosophers think that knowledge requires probability 1, and (3) certainly does not imply that, for any theoretical truth, it is possible that someone at some time will assign probability 1 to it. Second, neither (2) nor (3) ensures that if agents assign probability 1 to some truth, they will not be in a Gettier situation with respect to the given sentence, and thus will not still fail to know it. That our antirealism does not entail (8) might itself be thought to be a problem, for many antirealists seem to conceive (8) as a kind of adequacy condition for theories of truth. However, as is argued in Douven [2007a], and as we shall briefly see below (section 5), they are misguided on this point.

[8] Fitch 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 [2007].

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

4. Intuitive Correctness. In section 1 we noted that if (1) were adopted as our definition of truth for evidence sentences, 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 (2) instead of (1) is no guarantee that our theory satisfies this condition; it just prevents the theory from failing to satisfy it too obviously. Unfortunately, it is hard to determine whether our theory indeed satisfies this condition. This has to do with the fact that 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. Given that our definition of rationality is relatively weak, this will not help us to answer the question whether for any, or at least for most, sentences we deem pretheoretically 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 to some extent 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 will depend on the precise nature of the strengthening, and here it is regrettable that neither the aforementioned authors nor we have any concrete proposals for a strengthening on offer.

Does our inability to say anything definitive about our theory in connection with the present adequacy condition yield a point in favor of realist truth (which quite clearly does appear to satisfy the condition)? Perhaps not quite. For at least from a realist perspective it may seem likely that at least *extensionally* truth as defined in section 1 (given *any* of the options presented in subsection 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 explain.

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]. Very 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 first 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 (2) 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.[10] Thus the realist could not possibly think that our theory is anemic.

But our statement of Gaifman and Snir's result was rough, as we said, and this is so for various reasons. The ones most relevant to present concerns are, first, that the result has been proved only for a particular type of language, namely, one that consists of the language of first-order arithmetic with finitely many empirical predicates and function symbols added to it, and second, that the evidence sentences are assumed to *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.[11]

As to the former, it should be noticed that the question whether similar results hold for more expressive languages is still open. Currently no stronger results than that of Gaifman and Snir are available, but it is certainly hoped that the kind of convergence they prove to exist holds more—and even quite—generally. For readers who do not share this hope we mention that, by way of alternative response, we could restrict our truth definition to the language Gaifman and Snir consider; that would still be progress, given that at present for no even minimally representative fragment of a natural language does there exist an antirealist definition of truth.

As to the latter, until relatively recently most philosophers would have said that the assumption of separation 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.[12] (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.) While this thesis 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 EET.[13] 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. Exactly what the scientific realists' arguments against EET are need not detain us here; see, for instance, Douven [2007b] 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 as regards EET, by far most semantic realists are also *scientific* realists.[14] 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.[15]

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. Similarly, 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—as may already be

[10] This will be so regardless of which of the options considered in subsection 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).

[11] Actually the assumption is a bit weaker, namely, that the evidence are "almost everywhere separating," meaning that they separate the models in a class of models of measure 1; see Gaifman and Snir [1982:510].

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

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

[14] In 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.

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

clear from subsection 1.5—on which of the options is taken for extending the partial truth definitions (2) and (3) to the rest of the language. For example, if we let the whole truth definition consist of (2) plus (7), then it leaves open the possibility that a disjunction is true without either disjunct being true: conditional on more and more evidence sentences one may be increasingly certain that a given disjunction is true without ever thinking of either disjunct that it is more likely true than not. Obviously this is not so if we choose to define truth for non-atomic theoretical sentences recursively by means of the Strong Kleene Scheme. An important thing to note here is that, again, extensionally there may be no difference between the various theories of truth. Although it *need* not hold that the probability of a disjunction goes to 1 in the limit only if the probability of at least one of its disjuncts goes to 1 too, it *may* hold. And 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, there is no difference indeed between the various theories.

Finally, it will not have been missed that our definition of truth for atomic theoretical sentences assumes that the true evidence sentences come to be known to the community of rational agents in a determinate order. But—one may wonder—if they had become known 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—but certainly not necessary—that different orderings of the evidence sentences lead to different truth values of the atomic theoretic 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 known—which they need not be—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 any such possibility will materialize is zero). It thus seems that from the perspective of the realist the possibility that antirealist truth, as defined according to our proposal, may be order-dependent is not one to be taken seriously (and from the perspective of the antirealist it should appear to be little more than a matter of course, 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, either recursively or directly (in case we opt for (7)), in terms of people's probabilities for them conditional on more and more true evidence sentences. One may still wonder, however, whether it serves the purposes that have motivated antirealists.

The perhaps best-known motivation is of a meaning-theoretic nature and has been argued for most forcefully by Dummett.[16] 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. Another motivation for antirealism comes from the anti-skeptical sentiment that a theory that is ideal in that it entails all and only true evidence sentences and satisfies all theoretical virtues (such as simplicity and explanatory force) cannot possibly be false. At least it is not obvious that this follows from our theory of truth.[17]

But, as for the former motivation, it is important to note that it 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.[18] Once this is recognized, it is easy to show that any theory of truth entails that knowledge of sentence-meaning is fully manifestable *and* that no ideal theory can be false if it entails the following:

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

where for the purposes of this debate 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 [2007a] for the arguments. And given any (in the present context) reasonable interpretation of the word "possible" in (10)—like "logically possible" or "metaphysically possible"—our theory does entail (10): Firstly, if an evidence sentence is true according to (2), 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.[19] Secondly, it is shown in the Appendix that, given any of the truth definitions presented in section 1 (that is, whether we use the Strong Kleene or the supervaluation scheme to extend (3) to the rest of the language, or rather generalize (3) to all theoretical sentences), if a sentence φ is true, then,

[16] See, for instance, Dummett [1976].

[17] Kvanvig [2006, Ch. 2] mentions materialism and theism as further possible motivations for antirealism. However, he admits that materialism does not seem to motivate antirealism if the former is construed as a substantive thesis, which commits materialists to the truth or approximate truth of contemporary physics. He thinks such a construal is problematic because of the so-called pessimistic meta-induction (Laudan [1981]), but he seems unaware that, especially in view of Kitcher's [1993, Ch. 5] and Psillos's [1999, Ch. 5] work on the pessimistic meta-induction, it has lost much of its skeptical force. As for the motivation from theism, Kvanvig admits (p. 49) that this is weak, given that it not only depends on the plausibility of theism but also on that of some rather specific theistic assumptions.

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

[19] It 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.

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 sentence come to be known to 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, it 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.

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 for now we shall call it, 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 (2) 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."[20]

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:

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

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 the conditions are sufficiently good for the appraisal of φ. 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), too.

[20] According 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 noncircular 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*.

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 (see, e.g., Tennant [1990] or van Dalen [1994]): 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 uncancelled assumptions; the rule which allows us to conclude $\Diamond\varphi$ from φ; and, finally, the following introduction and elimination rules for the subjunctive conditional, which should be uncontroversial:

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

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

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

(As Wright [2000:342 n] notes, the application of the necessitation rule in the last step seems superfluous, as it is already unsettling enough that the epistemically ideal conditions should 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 true, then we would have

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

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? We think it does. For while (2) *almost* has the form of (12), it is restricted to elements of $\mathcal{E}$.[21] And it seems that 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 $\mathrm{Tr}(Q_\varphi) \leftrightarrow (Q_{Q_\varphi} \mathrel{\Box\!\!\rightarrow} \mathrm{B}\,Q_\varphi)$ (or $\mathrm{Tr}(Q_\varphi) \leftrightarrow (Q_\varphi \mathrel{\Box\!\!\rightarrow} \mathrm{B}\,Q_\varphi)$) a valid instantiation of (2).[22] As a result, the Plantinga–Wright argument does not apply to (2).

It could still apply to our *theory*, of course, if that could somehow be represented as being of the form (11) or (12). We are unable to see how it could, but we are happy to formulate this as a challenge to anyone who doubts that our theory dodges the above problem.

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 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

[21] Or if we can define generally the sufficiently good conditions for the appraisal of the elements of $\mathcal{E}$, (2) has the even simpler form of (11), again restricted to evidence sentences of course.

[22] Nor 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.

sentences could amount to, it is even harder to imagine that such conditions could be generally characterized.[23,24]

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 somehow 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 (2). The sentence "Electrons have negative charge" can be true without there being sufficiently good conditions for its appraisal; its truth (if it is true) requires that the limit of the probability any rational inquirer assigns to it conditional on more and more true evidence sentences equals 1.

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 of the fact that 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 logically incoherent.

We have been concerned with developing a Peircean conception of truth. It emerged that Peirce's antirealist credo applied to truth only carries us so far. If it is to be developed into a full-fledged theory of truth satisfying minimal adequacy conditions, it needs to be supplemented by conceptual ingredients—like, most notably, the key concepts of Bayesian epistemology—that exceed the idea in its slogan-like form.

[23] 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.

[24] 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). 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.

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 difficulties marring antirealist conceptions of truth. Not every antirealist theory of truth is equally vulnerable to all such objections that have been articulated in the literature. In fact we have been aiming at more. We hold out the hope that by skillfully combining tenets of several antirealist conceptions of truth (like Putnam's and Peirce's) all the technical difficulties that have been levelled against antirealist truth can be circumvented.

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.

Appendix

In section 5 it was claimed that, given any of the truth definitions proposed in section 1, if a sentence is true, then the probability an agent assigns to the sentence will converge to 1 in the limit, as more and more evidence sentences come to be known to the community of inquirers. For the theory that results if truth for all theoretical sentences is defined by generalizing (3) to all such sentences, this relationship between truth and probability 1 "in the limit" holds by definition, of course. For the other two theories, which use the Strong Kleene Scheme and the supervaluation scheme, respectively, it may be less obvious that the relationship holds. However, given that rational agents update by means of Bayes's rule, the relationship's holding for the former theory is an immediate corollary to Proposition A.1, and its holding for the latter is an immediate corollary to Proposition A.2.

The proofs of the following propositions make use of the notion of the *rank* of a sentence, where the rank $r(\varphi)$ of a sentence $\varphi \in \mathcal{L}$ is defined as follows:

- $r(\varphi) = 0$ if φ is atomic;
- $r(\neg\varphi) = r(\varphi) + 1$;
- $r(\varphi \vee \psi) = \max(r(\varphi), r(\psi)) + 1$;
- $r(\exists x\varphi(x)) = r(\varphi(d)) + 1$, with $d \in \mathcal{D}$.

Proposition A.1 *For all $\varphi \in \mathcal{L}$ and all rational agents $i \in I$, if $V_{SK}(\varphi) = 1$, then $\lim_{t\to\infty} \Pr_i(\varphi \mid \bigwedge K_t) = 1$, and if $V_{SK}(\varphi) = 0$, then $\lim_{t\to\infty} \Pr_i(\varphi \mid \bigwedge K_t) = 0$.*

Proof: The proof is by induction over the rank of sentences. The basis step follows immediately from (2), (3), and (4). Thus assume that the proposition holds for all sentences of rank $\leqslant n$. It is then to be shown that it holds for all sentences φ of rank $n + 1$. There are several cases to be considered. For ease of presentation, we will leave quantification over the rational agents implicit, both in this proof and in the proof of Proposition A.2.

- $\varphi = \neg\psi$, for some sentence ψ: If $V_{SK}(\varphi) = 1$, then $V_{SK}(\psi) = 0$. So, by the induction hypothesis, $\lim_{t\to\infty} \Pr_i(\psi \mid \bigwedge K_t) = 0$. So, by probability theory, $\lim_{t\to\infty} \Pr_i(\neg\psi \mid \bigwedge K_t) = 1$. And so, $\lim_{t\to\infty} \Pr_i(\varphi \mid \bigwedge K_t) = 1$. One proves in a symmetrical fashion that $\lim_{t\to\infty} \Pr_i(\varphi \mid \bigwedge K_t) = 0$ if $V_{SK}(\varphi) = 0$.

- $\varphi = \psi \vee \chi$, for some sentences ψ and χ: If $V_{SK}(\varphi) = 1$, then $V_{SK}(\psi) = 1$ or $V_{SK}(\chi) = 1$ (or both). If the first, then, by the induction hypothesis, $\lim_{t\to\infty} \Pr_i(\psi \mid \bigwedge K_t) = 1$. It then follows by probability theory that $\lim_{t\to\infty} \Pr_i(\psi \vee \chi \mid \bigwedge K_t) = 1$. Similarly if $V_{SK}(\chi) = 1$. Hence, $\lim_{t\to\infty} \Pr_i(\varphi \mid \bigwedge K_t) = 1$. If, on the other hand, $V_{SK}(\varphi) = 0$, then both $V_{SK}(\psi) = 0$ and $V_{SK}(\chi) = 0$. Then, by the induction hypothesis, both $\lim_{t\to\infty} \Pr_i(\psi \mid \bigwedge K_t) = 0$ and $\lim_{t\to\infty} \Pr_i(\chi \mid \bigwedge K_t) = 0$. By probability theory it then follows that $\lim_{t\to\infty} \Pr_i(\psi \vee \chi \mid \bigwedge K_t) = 0$ and thus that $\lim_{t\to\infty} \Pr_i(\varphi \mid \bigwedge K_t) = 0$.

- $\varphi = \exists x\psi(x)$, for some open formula $\psi(x)$: If $V_{SK}(\varphi) = 1$, then, for some $d_k \in \mathcal{D}$, $V_{SK}(\psi(d_k)) = 1$. Thus, by the induction hypothesis, $\lim_{t\to\infty} \Pr_i(\psi(d_k) \mid \bigwedge K_t) = 1$ for some d_k. So, by probability theory, $\lim_{t\to\infty} \Pr_i(\exists x\psi(x) \mid \bigwedge K_t) = 1$. And so, $\lim_{t\to\infty} \Pr_i(\varphi \mid \bigwedge K_t) = 1$. If, on the other hand, $V_{SK}(\varphi) = 0$, then, for all $d_k \in \mathcal{D}$, $V_{SK}(\psi(d_k)) = 0$. Then, by the induction hypothesis, $\lim_{t\to\infty} \Pr_i(\psi(d_k) \mid \bigwedge K_t) = 0$ for all d_k. By probability theory it follows from this that $\lim_{t\to\infty} \lim_{n\to\infty} \Pr_i(\bigvee_{k=0}^{n} \psi(d_k) \mid \bigwedge K_t) = 0$. Again by probability theory, in particular by axiom (A4), this yields $\lim_{t\to\infty} \Pr_i(\exists x\psi(x) \mid \bigwedge K_t) = 0$. And so, $\lim_{t\to\infty} \Pr_i(\varphi \mid \bigwedge K_t) = 0$. ⊠

Proposition A.2 *For all $\varphi \in \mathcal{L}$ and all rational agents $i \in I$, if $V_{SV}(\varphi) = 1$, then $\lim_{t\to\infty} \Pr_i(\varphi \mid \bigwedge K_t) = 1$, and if $V_{SV}(\varphi) = 0$, then $\lim_{t\to\infty} \Pr_i(\varphi \mid \bigwedge K_t) = 0$.*

Proof: This proof follows the same strategy as the proof of Proposition A.1. Again the basis step follows directly from (2), (3), and (4), and again we assume that the proposition holds for all sentences of rank $\leqslant n$. Then the proof for the case where $\varphi = \neg\psi$ for some ψ is essentially the same as in the previous proof. This leaves the following cases to be considered:

- $\varphi = \psi \vee \chi$, for some ψ and χ: If $V_{SV}(\varphi) = 1$, then we have these subcases (it is easy to see that in the remaining subcases $V_{SV}(\psi \vee \chi) \neq 1$):

  - $V_{SV}(\psi) = 1$ or $V_{SV}(\chi) = 1$ (or both): In this subcase the proof proceeds basically as in the corresponding clause of the previous proof.
  - $V_{SV}(\psi)$ and $V_{SV}(\chi)$ are both undefined: Given that, nevertheless, $V_{SV}(\psi \vee \chi) = 1$, $\psi \vee \chi$ must be true on all possible assignments of truth values to ψ and χ. Hence $\psi \vee \chi$ must be a tautology. But then it must have been assigned probability 1 from the start. Hence, it must be that $\lim_{t\to\infty} \Pr_i(\psi \vee \chi \mid \bigwedge K_t) = 1$. And thus $\lim_{t\to\infty} \Pr_i(\varphi \mid \bigwedge K_t) = 1$.

  If $V_{SV}(\varphi) = 0$, then it must be that $V_{SV}(\psi) = V_{SV}(\chi) = 0$. Then, by the induction hypothesis, $\lim_{t\to\infty} \Pr_i(\psi \mid \bigwedge K_t) = \lim_{t\to\infty} \Pr_i(\chi \mid \bigwedge K_t) = 0$. And from this it follows that $\lim_{t\to\infty} \Pr_i(\psi \vee \chi \mid \bigwedge K_t) = 0$ and thus that $\lim_{t\to\infty} \Pr_i(\varphi \mid \bigwedge K_t) = 0$.

- $\varphi = \exists x\psi(x)$, for some $\psi(x)$: If $V_{SV}(\varphi) = 1$, then there are again two subcases to consider. If for some $d_k \in \mathcal{D}$, $V_{SV}(\psi(d_k)) = 1$, then the proof proceeds as in the corresponding clause of the proof of Proposition A.1. In the second subcase, in which, for all $d_k \in \mathcal{D}$, $V_{SV}(\psi(d_k))$ is undefined, we derive again the conclusion that $\exists x \psi(x)$ is a tautology, so that $\lim_{t\to\infty} \Pr_i(\exists x \psi(x) \mid \bigwedge K_t) = 1$ and hence $\lim_{t\to\infty} \Pr_i(\varphi \mid \bigwedge K_t) = 1$. If, on the other hand, $V_{SV}(\varphi) = 0$, then it must be that $V_{SV}(\psi(d_k)) = 0$ for all $d_k \in \mathcal{D}$. Then we again have that $\lim_{t\to\infty} \lim_{n\to\infty} \Pr_i(\bigvee_{k=0}^{n} \psi(d_k) \mid \bigwedge K_t) = 0$, which, by axiom (A4), gives us $\lim_{t\to\infty} \Pr_i(\exists x \psi(x) \mid \bigwedge K_t) = 0$. Thus, finally, $\lim_{t\to\infty} \Pr_i(\varphi \mid \bigwedge K_t) = 0$. ⊠

References

Achinstein, P. [2001] *The Book of Evidence*, Oxford: Oxford University Press.

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.

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. [2006] "Assertion, Knowledge, and Rational Credibility," *Philosophical Review* 115:449–485.

Douven, I. [2007a] "Fitch's Paradox and Probabilistic Antirealism," *Studia Logica*, in press.

Douven, I. [2007b] "Underdetermination," in S. Psillos and M. Curd (eds.) *The Routledge Companion to the Philosophy of Science*, London: Routledge, 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.

Foley, R. [1992] "The Epistemology of Belief and the Epistemology of Degrees of Belief," *American Philosophical Quarterly* 29:111–124.

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.

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. [1993] *The Advancement of Science*, Oxford: Oxford University Press.

Kitcher, P. [2001] "Real Realism: The Galilean Strategy," *Philosophical Review* 110:151–197.

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, H. [1970] "Conjunctivitis," in M. Swain (ed.) *Induction, Acceptance and Rational Belief*, Dordrecht: Reidel, pp. 55–82.

Laudan, L. [1981] "A Confutation of Convergent Realism," *Philosophy of Science* 48:19–49.

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

Maher, P. [1993] *Betting on Theories*, Cambridge: Cambridge University Press.

Peirce, C. S. [1878/1998] "How to Make Our Ideas Clear," in E. C. Moore (ed.) *Charles S. Peirce: The Essential Writings*, Amherst NY: Prometheus Books, 1998, pp. 137–157.

Plantinga, A. [1982] "How to Be an Anti-Realist," *Proceedings and Addresses of the American Philosophical Association* 56:47–70.

Psillos, S. [1999] *Scientific Realism: How Science Tracks Truth*, London: Routledge.

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. [2007] "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. [1980] *The Scientific Image*, Oxford: Clarendon Press.

Visser, A. [1989] "Semantics and the Liar Paradox," in D. Gabbay and F. Guenthner (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, 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. [1992] *Truth and Objectivity*, Cambridge MA: Harvard University Press.

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