Routledge
Taylor & Francis Group

The pessimistic induction, the flight to reference and the metaphysical zoo

Michael A. Bishop

Department of Philosophy, Northern Illinois University, DeKalb, Illinois, USA

Abstract   *Scientific realism says of our best scientific theories that (1) most of their important posits exist and (2) most of their central claims are approximately true. Antirealists sometimes offer the pessimistic induction in reply: since (1) and (2) are false about past successful theories, they are probably false about our own best theories too. The contemporary debate about this argument has turned (and become stuck) on the question, Do the central terms of successful scientific theories refer? For example, Larry Laudan offers a list of successful theories that employed central terms that failed to refer, and Philip Kitcher replies with a view about reference in which the central terms of such theories did sometimes refer. This article attempts to break this stalemate by proposing a direct version of the pessimistic induction, one that makes no explicit appeal to a substantive notion or theory of reference. While it is premature to say that this argument succeeds in showing that realism is probably false, the direct pessimistic induction is not subject to any kind of reference-based objection that might cripple a weaker, indirect version of the argument. Any attempt to trounce the direct pessimistic induction with a theory of reference fails.*

Scientific realists and antirealists look at the history of science and draw contrary conclusions. Realists tend to focus on the successes of science and argue that (1) most of the important posits of our best scientific theories exist and (2) most of the central claims of our best scientific theories are true or approximately true. Antirealists tend to focus on the failures of science and argue that both of these claims are false. Perhaps the best-known antirealist argument is the *pessimistic induction*: from the fact that (1) and (2) are false about successful theories in the past, we should suppose that they are false about our own best theories as well.

The core issues separating realists and antirealists are about truth (or approximate truth) and ontology. Are the claims of our best theories (approximately) *true*? Do most of the posits of our best theories *exist*? A flight-to-reference argument tries to settle such issues by appeal to theories of reference (Bishop & Stich, 1998). Larry Laudan's classic presentation of the pessimistic induction is a flight-to-reference argument. The linchpin of Laudan's argument is the claim that the history of science shows that successful theories "contained central terms that (we now believe) were nonreferring" (Laudan, 1984a, p. 121). The most prominent realist responses to the argument are also flight-to-reference arguments (Putnam, 1978; Boyd, 1984; Kitcher, 1993; Leplin, 1997). For example, Philip Kitcher's important defence of realism against the pessimistic induction tries to show that many of the central terms of successful obsolete

theories did, on occasion, refer (Kitcher, 1993, p. 149). I will argue that it has been a serious mistake for both realists and antirealists to frame this debate in terms of reference. My aim in this article is to free this debate from the murky grips of the philosophy of language by proposing a version of the pessimistic induction that cannot be undermined by any appeal to a substantive theory of reference. I do not claim that the pessimistic induction succeeds in showing that realism is probably false. The success of the pessimistic induction depends on a host of historical claims that I can't begin to defend here. My purpose is to offer a framework for making the pessimist's case that can resist flight-to-reference arguments.

The article will proceed as follows. In Section 1, I begin with some preliminaries about semantics and ontology. Nothing in this section is particularly controversial, but getting straight on these matters will be essential to the argument. In Section 2, I propose a *direct* version of the pessimistic induction, one that does not employ a substantive theory or notion of reference. In Section 3, I argue that the standard realist response to the pessimistic induction, a reference-based response, leaves untouched the direct pessimist's argument. The point here is *not* that the direct pessimistic induction succeeds. Rather, the point is that the direct argument is not subject to any kind of reference-based objection that might cripple a weaker, indirect version of the argument. In Section 4, I consider and reply to three natural responses to the direct pessimistic induction. The fifth and final section briefly recaps and concludes.

1. Preliminaries: semantics and ontology

Semantics is concerned with language, in particular, with meaning, reference and truth. Ontology is concerned with existence, in particular, with what there is. There are many debates about the relationship between semantics and ontology, but most realists take them to be quite different sorts of endeavours. What exists in the parts of the world that don't have linguistic creatures in them does not depend on language. For the realist, and for the clear-headed antirealist, the existence or non-existence of oxygen, quarks, light waves, ether, etc., does not depend on the contingent fact that humans exist or have minds or speak language or have the particular languages that we do. Our theories of chemistry and physics do not stand or fall depending on the outcome of a philosophical dispute about the nature of language. Theories of meaning and reference are silent about such matters.

So far, so good. Realists will typically agree to (indeed, insist upon) something like the following claims.

1. The realism–antirealism debate is at bottom a debate about ontology, about what there is. If the posits of our best theories don't exist, antirealism wins the day; if the posits of our best theories do exist and if what they say about those posits is roughly true, then realism wins the day.
2. Ontology does not depend upon semantics. What exists in the parts of the world that don't have linguistic creatures in them does not depend in any way on language.

And yet realists will also often insist, at least implicitly, on a third claim, particularly when responding to the pessimistic induction.

3. Reference is essential to issues of ontology. The way to figure out whether a posit exists is to implicitly or explicitly invoke reference.

Prima facie, (2) and (3) are a contradiction. Claim (2) says semantics is not essential to ontology; claim (3) says semantics is essential to ontology. So what's going on? I want to suggest that there is a deep ambiguity in our notion of reference. (This is not a novel claim; many philosophers have identified an ambiguity similar to the one I describe here.[1]) Sometimes, we employ an ontological conception of reference, one in which "T" refers exactly if T (or Ts) exist(s). This notion of reference is embodied by a deflationary principle:

(O) "T" refers iff T (or Ts) exist.[2]

In schema (O), the same expression is substituted in both occurrences of "T". And (O) is defined only for expressions a speaker understands, so (O) does not have the consequence that anything that exists must have a term that designates it. Sometimes, however, we employ a substantive conception of reference—"T" refers exactly if "T" stands in a substantive reference-relationship to something. In other words:

(S) "T" successfully refers iff (Ex) ("T" Ref x).

A *substantive* theory of reference is an account of the substantive relation (Ref) that holds between an expression and what it denotes. Examples include causal-historical theories (Kripke, 1972; Putnam, 1975; Devitt, 1981), description theories (Russell, 1919; Searle, 1958; Lewis, 1970, 1972) or mixed causal-description theories (Kitcher, 1978, 1993; Evans, 1983). It's possible for one to adopt a substantive theory of reference that makes the substantive and deflationary conceptions of reference extensionally equivalent. But that's not the way it usually works. Substantive theories of reference typically define notions of reference that do not make all substitution instances of (O) true. The reason is that substantive theories of reference typically have an interpretive function, whereas this is not the case with the deflationary notion of reference. Two examples will help clarify this point.

Suppose I believe a theory of light that includes a posit designated by the term "Gremlin". Part of my theory says that Gremlins are conscious and are responsible for both light phenomena and the production of beer in Ecuador. Suppose I point at the open window and say, "Please draw the shade because the Gremlins are hurting my eyes."[3] We might reasonably say that my utterance of the expression "Gremlins" successfully refers in the sense given by (S): there is something to which my utterance stands in the appropriate reference-relationship, namely light. But there is another sense, given by (O), in which we would want to say that "Gremlins" doesn't refer—that's because Gremlins don't really exist. So it's possible for an expression token not to refer in the deflationary sense and to refer in the substantive sense. We can say that "Gremlins" on this occasion s-refers but does not o-refer.

Consider another example. Philip Kitcher's theory of reference yields the result that "dephlogisticated air" when used by Priestley and "light wave" when used by Maxwell do on occasion successfully refer. What notion of reference is at work here? Kitcher never commits himself to the actual *existence* of the posits of these obsolete theories, i.e. dephlogisticated air and light waves. Rather, Kitcher argues that "light wave" refers on occasion to electromagnetic radiation and "dephlogisticated air" refers on occasion to oxygen. This strongly suggests that Kitcher is employing a substantive, rather than a deflationary, conception of reference. That is, Kitcher is arguing that these obsolete expressions s-refer, but he is not committed to the claim that they o-refer.

Notice that in both of these cases, the substantive conception of reference serves to

allow us to interpret people's utterances. But from the fact that these expressions successfully s-refer, it does not follow that they successfully o-refer. For our purposes, there is no need to argue that one of these conceptions of reference is more legitimate, more important or more central than the other one. I would contend that both are important and perfectly respectable. What is not legitimate, however, is to employ a notion of reference that equivocates between o-reference and s-reference.

Now we are in a position to understand the apparent contradiction between (2) and (3). How can reference be essential and not essential to ontology? We need to be a bit careful here. On both conceptions of reference, if an expression successfully refers, that means that a referent *exists*. So in this minimal sense, both conceptions of successful reference bring with them ontological commitments. But the ontological issue that divides scientific realists and scientific antirealists isn't whether *something or other* exists. It's whether *the posits of our best theories* exist. With respect to this ontological issue, s-reference is not essential, but o-reference is (trivially) essential. Both notions of reference bring an ontological commitment, but only o-reference brings a *specific* ontological commitment. For example, suppose we're interested in the ontological question of whether quarks exist and we have adopted a substantive theory of reference that yields the result that "quark" s-refers. It follows that the expression "quark" refers to something. But that's not enough to conclude that *quarks* exist, any more than the fact that "Gremlins" s-refers allows us to conclude that Gremlins exist. Now, if one could establish that "quark" *o-refers*, then one could establish that quarks exist. That's because to establish that "quark" o-refers *just is* to establish that quarks exist. Since the issue of whether "T" o-refers is identical to the issue of whether T(s) exist(s), then of course o-reference is (trivially) relevant to the ontological issues separating realists and antirealists.

2. The direct formulation of the pessimistic induction

Now that we have the preliminaries out of the way, we can proceed to the direct formulation of the pessimistic induction. What distinguishes direct formulation from the indirect formulation of the argument is that the direct formulation does not appeal to *s-reference*. Since it is about whether the posits of past and present theories *exist*, it trivially involves *o-reference*. The centrepiece of the direct pessimist's argument is a robust evidential base consisting of past successful theories whose important posits do not exist, and whose central claims are neither true nor approximately true.[4] From this evidence, the direct pessimist infers by enumerative induction that probably (1′) most of the important posits of our best scientific theories do not exist, and (2′) most of the central claims of our best scientific theories are neither true nor approximately true. In order to build this evidential base, the direct pessimist will try to get some purchase on what scientific posits *do* exist and what scientific claims *are* true. In this way she can show that many obsolete but successful theories got these matters wrong. But there's the rub: the antirealist doesn't think she (or anyone else) has any justified beliefs about what scientific posits exist or what scientific theories are true. For this reason, the direct pessimist begins her argument with a fiction. She supposes that the posits of our best current scientific theories exist and that the central explanatory claims of those theories are true. (In place of this mouthful, I will sometimes follow the useful but inaccurate convention of saying our best theories are true.) Let's explore the fiction in a bit more detail.

2.1. The fiction

The pessimist pretends that what our best theories say about the world is true. (D) tells us what the pessimist means by this.

(D) "S" is true iff S.

In schema (D), the same sentence is substituted for both occurrences of "S". (D) is defined only for sentences that a speaker understands (Field, 1994a, b). As a result, (D) can be applied without recourse to a *substantive* theory of reference. Substitution instances of (D) are mind-numbingly obvious. To say that "The speed of light is constant" is true is just to say that the speed of light is constant. (D) is the only general hypothesis about truth that the direct pessimist needs to get her argument started. It is worth noting that nothing here commits the direct pessimist to the bold claim that something like (D) exhausts our notion of truth (Horwich, 1990; Field, 1994a, b). She requires only the weaker claim that any theory of truth must make all appropriate substitution instances of (D) true.

The second part of the pessimist's fiction is the assumption that the posits of our best current theories exist. How are we to understand this? Here is a deflationary schema about posits designated by singular or general expressions.

(O′) The posit of a theory designated by term "T" exists iff T (or Ts) exist.

(O′) is a variant of schema (O) (introduced above) and defines a notion of o-reference for scientific expressions. The direct pessimist will use this schema to interpret the realist claim that the posits of our theories exist. For example, our best theories contain the expression "quark". The relevant substitution instance of schema (O′) would be: the posit designated by "quark" exists just in case quarks exist. (Instead of this, I will continue to say that "quark" o-refers.) It is worth noting that the direct pessimist is not committed to the claim that (O′), (O) or something like these schemas, exhaust our notion of reference or designation. (For such a view, see Horwich, 1990, pp. 110–125.) The direct pessimist is free to accept plenty of substantive theories about the nature of reference which go well beyond a minimalist schema like (O) or (O′).

The direct pessimist is now in a position to make judgements about whether the posits of obsolete theories exist without appeal to s-reference. Suppose the issue is whether or not the ether (the posit of obsolete theories of light) exists. *Assuming our current theory of light is true*, we can draw up a comprehensive list of all the posits that explain the phenomenon of light. For each member of the list, we can ask, "Is that identical to the ether?" If the answer to all our questions is "No" then the ether doesn't exist. If the answer to one of our questions is "Yes" then it does. (Notice two points about this procedure. First, it works only for scientific posits. Bottle caps exist even though they aren't identical to any posit of our best current scientific theories. Second, this procedure is not equivalent to one that simply checks whether our current theories retain the *word* "ether". The procedure aims to discover whether our current theories posit the ether, regardless of what it might be called.)

Some realists are unlikely to be convinced by this procedure. Perhaps they will insist that while this procedure avoids the *vocabulary* of s-reference, it nonetheless smuggles in a substantive theory of reference. This is a crucial point. I will postpone a discussion of this issue until Section 4.1. For now, I will continue with the direct formulation of the pessimistic induction on the assumption that we can make at least a good number of ontological judgements without a substantive theory of reference.

2.2. Laudan's list

Armed with the pretense that our best theories are true, the direct pessimist sets out to construct an evidential base of successful theories whose posits don't exist and whose central claims aren't true or approximately true. Laudan (1984a) is strong on this point. Laudan's list includes wave theories of light, particle theories of light, the humoral theory of medicine, the effluvial theory of static electricity, catastrophist geology, the phlogiston theory of chemistry, the caloric theory of heat, the vibratory theory of heat, the vital force theories of physiology, the theory of circular inertia, and theories of spontaneous generation. Laudan points out, "This list, which could be extended ad nauseam, involves in every instance a theory that was once successful and well confirmed, but which contained central terms that (we now believe) were nonreferring" (Laudan, 1984a, p. 121). Laudan's explicit desideratum for choosing these theories is that their central terms did not refer. The direct pessimist likes the list, but she eschews all talk of reference. The important feature of the theories on Laudan's list is that *their central posits do not exist*. (Of course, this is trivially equivalent to the claim that the central terms of these theories do not o-refer. But what's important for the direct pessimist is that her argument not employ any notion of s-reference.) To see how the direct pessimist would handle these theories without adverting to a substantive theory or notion of reference, let's consider an example.

Wave theories of light posited the existence of a pervasive luminiferous ether, and light was supposed to consist of waves propagated through this ether. The central explanatory posit of this successful theory was the *light wave*. Do light waves exist? The direct pessimist returns to the fiction. Let's assume the posits of our best current theories exist, in accordance with schema (O). Do our theories leave a place in the world for light waves? No. One reason wave theories of light were overthrown is that physicists discovered that light didn't really consist of *waves*; it consists of photons. Why believe this claim? Certainly not on philosophical grounds. Empirical arguments (e.g. arguments from blackbody radiation and the photoelectric effect) convinced physicists that the central posits of wave theories don't exist. Here's a vivid way to put the point. There are possible worlds in which the posits of Maxwell's theory really exist. But our world isn't one of them. And we know this because of the overwhelming empirical arguments physicists have offered over the past century or so.[5]

I can imagine a radical relativist of some sort disputing this point. But I can't imagine a *realist* insisting that we reject all the empirical arguments of present-day physicists concerning the *ontological* failures of 19th-century theories of light. Realists might well argue that the expression "light wave" in Maxwell's theory s-refers to photons. But to say that light waves (the posits of 19th-century theories of light) exist is a very different claim. Let's say one believes in a *metaphysical zoo* when one holds that the central posits of past and present successful scientific theories all really exist—including those posits that present-day scientists (believe they) have discredited on the basis of powerful and (as of yet) undefeated empirical arguments. Some inhabitants of the metaphysical zoo might include Bohr's atoms, dephlogisticated air, ether waves, etc. I am aware of no present-day realist who has said in print that the atoms of Bohr's early theories or Priestley's dephlogisticated air or Maxwell's ether waves really *exist*.[6] Instead, realists say that Bohr referred to atoms with the term "atom", Priestley referred to oxygen when he used the expression "dephlogisticated air", etc. If scientific realists embrace the metaphysical zoo, they've done an extraordinary job keeping this a secret. Until and unless realists break down and admit to believing in the metaphysical zoo (in

believing, for example, that we live in a universe in which Maxwell's posits really exist), I will assume realists will accept this step in the direct pessimistic induction. (For more discussion of this view, see Section 3.)

So the central explanatory posit of the wave theory of light doesn't exist. Now the pessimist asks whether the central claims of the wave theory of light are true. For example, is "Light consists of light waves" true? Since light waves don't exist, it can't be true that light consists of them. Therefore, given (D), the sentence can't be true either. This argument can be generalized. No statement of the wave theory that contains the expression "light wave" can be true. So the direct pessimist is right about this theory. Its central explanatory claims (in so far as they appeal to light waves) are not true.[7]

At this point, the realist might object that this standard is too high. The realist needs to show only that the central claims of the theory are *approximately* true, and this is perfectly consistent with their claims being strictly false. How should the pessimist deal with approximate truth? With the following schema.

> (A) If the expression "a" in sentence "Fa" designates a posit that does not exist (in accordance with schema [O′]), then "Fa" is not approximately true.[8]

Schema (A) has a lot of intuitive appeal. To say that a sentence is approximately true is to say that it *just missed* having been true about the world. But if the subject of a sentence doesn't exist, then the sentence isn't *about* the world (in any sense that could give comfort to the realist). So it can't have *just missed* having been true about the world. It missed by quite a lot. Given schema (A) and others like it, the pessimist can argue that since the central posits of the phlogiston theory and the wave theory don't exist, the central claims of those theories aren't approximately true. And so too for all the theories on Laudan's list whose central posits don't exist. (For a debate about whether realists can adopt a notion of approximate truth that rejects [A], see Hardin & Rosenberg, 1982 and the reply by Laudan, 1984b.)

My purpose here is not actually to construct an evidential base for the direct pessimistic induction. Rather, it is to sketch a strategy for constructing such a base without appeal to s-reference: for each successful, obsolete theory, show that its central posits don't exist with schema (O′), and then with schemas (D) and (A) show that the theory's central claims are neither true nor approximately true. In order to reach realist or antirealist conclusions about the theories on Laudan's list, there is no substitute for slogging through each theory and applying the relevant schemas.

2.3. The conclusion: taking back the fiction

Suppose the pessimist finds that all the theories on Laudan's list (and many other successful theories) were false, i.e. their central posits didn't exist and their central claims weren't true or approximately true. There is now a fund of evidence that should be acceptable to the realist. A large number of currently obsolete theories had the following properties: (a) they were pragmatically and predictively successful, (b) their central posits don't exist, and (c) their central claims are neither true nor approximately true. After having constructed her evidential base, the pessimist is ready to take back the fiction she granted the realist at the outset. Of the successful theories we currently believe, it is no longer acceptable to simply assume they are true. We need evidence. And the pessimist has a lot of evidence for thinking the realist is wrong—the fund of successful theories that posited non-existent entities and said false things about the world *even on the optimistic assumption that our current theories are true*. By enumerative

induction on the basis of this evidence, the direct pessimist argues that we should suppose our current theories posit non-existent entities and what they say about the world is neither true nor approximately true. The central tenets of scientific realism are false.

The realist might wonder whether it's that easy for the pessimist to take back her fiction. Some have suggested that realism is in some sense epistemically privileged because the *only* way to support antirealism is to presuppose realism (Leplin, 1997; Nagel, 1997). But at least in this case, that's false. The direct pessimistic induction, as I have so far articulated it, is no more problematic than any *reductio ad absurdum*. Indeed, the best way to understand the antirealist's project is to see the pessimistic induction as a sub-argument in a grand dilemma. Either realism is true or it's false. If realism is true, then (by the pessimistic induction) antirealism is true. And if realism is false, then antirealism is true. Either way, antirealism is true. So it's false that the scientific antirealist *must* assume that scientific realism is true in order to argue for her position; the direct pessimist need only assume that scientific realism is either true or false.

3. The problem with the realist's flight to reference

There are a number of moves available to the realist in response to the pessimistic induction. One is to winnow down the theories on Laudan's list. Realism is committed only to claiming that suitably *successful* theories are (approximately) true. Realists have argued that many theories on Laudan's list are false, but not successful enough to cause the realist any worries (e.g. Kitcher, 1993; Leplin, 1997). But as far as I know, no one has argued that this winnowing process will eliminate all the theories on Laudan's list. For the rest of the theories, realists typically offer a substantive account of reference that yields the conclusion that the central expressions of suitably successful obsolete theories did, as a matter of fact, refer (Putnam, 1978; Boyd, 1984; Kitcher, 1993; Leplin, 1997). This is a classic flight-to-reference argument—an argument that tries to resolve issues about truth and ontology (which, at bottom, is what the realism–antirealism debate is all about) by appeal to s-reference. While the details of these theories of reference are interesting and important, they really don't matter to the direct pessimist. She is happy to grant whatever semantic theory the realist wants to defend. However, the direct pessimist must keep s-reference and o-reference distinct. From the realist's substantive theory of reference, the direct pessimist will grant only claims of the form: *"T" successfully s-refers*. From such claims, it is illegitimate to conclude that T(s) exist(s).

The fundamental problem with the realist's flight-to-reference argument can be represented in a flow chart (Figure 1). Question (1) asks whether expressions like "light wave" or "dephlogisticated air" refer, according to the realist's theory of reference. If not, then the realist's theory of reference does not protect realism from Laudan's pessimistic induction (2). If the central terms of successful obsolete theories don't refer, then their central claims can't be true or approximately true. And antirealism wins the day. So such expressions must refer, if the realist's theory of reference is going to serve the realist's purposes. Question (3) asks whether the realist is willing to apply schema (O) to the obsolete expressions that refer, according to the realist's favoured theory of reference. (Recall that (O) says: "T" refers iff T (or Ts) exist(s).) Suppose (4) the realist is not willing to apply schema (O). Then the direct pessimist wins the day. The realist's claims about s-reference do not touch on the real ontological issue separating realism and antirealism. If the realist is willing to apply the schema (5), then the realist must
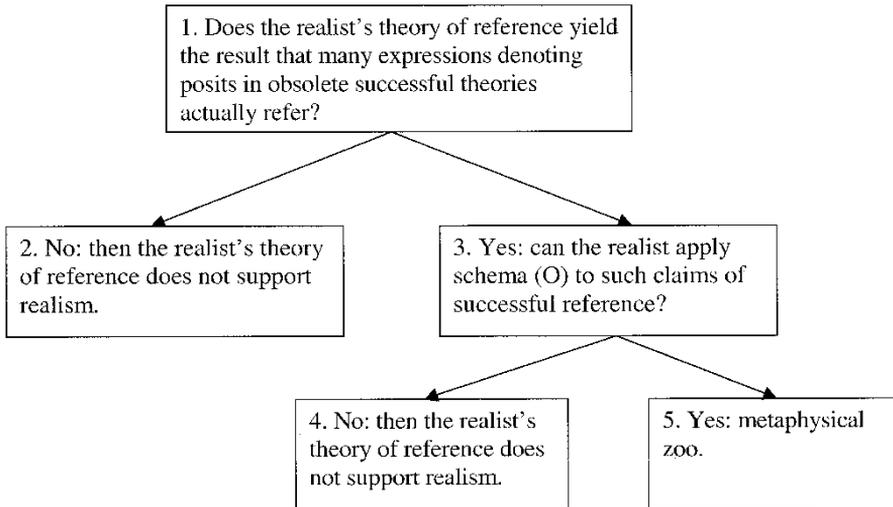
Figure 1. *The direct pessimist's flow chart.*

embrace the metaphysical zoo; the realist must admit, not just that the *expressions* for the obsolete posits of successful theories *refer*, but that those posits actually *exist*.

Let's suppose the realist avoids the metaphysical zoo and stops at (4). Suppose further that the realist argues that "light wave" referred to photons (or electromagnetic radiation) and "dephlogisticated air" referred to oxygen. If so, then it would appear that the realist can reasonably contend that, while the posits of these theories did not exist, the terms nonetheless s-refer to real things (photons, oxygen) and so the central claims of those theories could be true or approximately true. But it's important to see why this argument won't work. Consider this claim: *"Light wave" s-refers to photons*. If we take claims of this form to be ontologically committed to (in this case) photons, then the realist is not entitled to such claims. This point is essential. The direct pessimist argues that we have inductive evidence for thinking that the posits of our best theories, including oxygen and photons and all the rest, don't exist. This argument may be all wrong. But the realist can't defeat it armed only with a substantive theory of reference, a theory about language. The direct pessimist need not be shy about granting realist claims about language—*as long as they are only about language*. So the direct pessimist can grant that "dephlogisticated air" refers to *whatever we refer to with the expression "oxygen"*. She can grant that "light waves" refers to *whatever we refer to with the expression "photons"*. And so on. In other words, armed with a substantive theory of reference, the realist is entitled to claims like 1b and 2b, but not claims like 1a and 2a.

1a. "Light waves" s-refers to photons.
1b. "Light waves" and "photons" co-s-refer.

2a. "Dephlogisticated air" s-refers to oxygen.
2b. "Dephlogisticated air" and "oxygen" co-s-refer.

The claims to which the realist is entitled (like 1b and 2b), claims about the referential features of various linguistic expressions, imply nothing whatsoever about whether oxygen, photons, phlogiston, ether or any other posits exist. The direct formulation of

the pessimistic induction is immune from all attempts to undermine it via a substantive theory of reference. The direct pessimist has nothing to fear from flight-to-reference realists.

Let's suppose the realist bites the bullet (5): admit realism is committed to the metaphysical zoo, but insist that this is perfectly reasonable. To say that light waves (the posit of 19th-century theories of light) exist is just to say that electromagnetic radiation exists; to say that dephlogisticated air exists is just to say that oxygen exists; and so on. For example, when Priestley posited dephlogisticated air, it turns out he was positing oxygen under a different description. This is not simply the view, associated with causal-description theories of reference, that one can s-refer to something without knowing very much about it (e.g. my expression "Plato" can s-refer to Plato even though I know next to nothing about Plato). Rather, this is the view that the posit of Maxwell's theory he called "light waves" exists in the actual world—not in the non-actual world in which what is called "light" consists of waves in a pervasive ether. This view of scientific posits is quite counter-intuitive. Still, some might embrace it for the sake of scientific realism. There are two main objections to it.

The first objection to the metaphysical zoo is that it leads to an Orwellian view about scientific language. Consider that physicists offer powerful evidence-based arguments showing that these sorts of posits (ether, light waves, dephlogisticated air) do not exist. What would the philosopher committed to the metaphysical zoo have us say to our colleagues in the physics or chemistry or history departments when they give us detailed, evidence-based arguments for why dephlogisticated air doesn't exist? Presumably, we're supposed to tell them that they (and just about everyone else who has participated in or commented on the debate) don't properly understand the expression "dephlogisticated air". They're really talking about oxygen, though they don't know it. Of course, Priestley didn't know what he was talking about either when he used the expression. Maxwell didn't know what he was talking about when he used the expression "light wave". And on the assumption that "light particle" (the expression for the central posit of particle theories of light) also refers to electromagnetic radiation, nobody has recognized that the wave theorists and particle theorists were really positing the same thing the whole time—electromagnetic radiation. The centuries-long debate over particle and wave theories of light? Mostly the result of semantic confusion. The great physicists who participated in the wave–particle debate (Newton, Huygens, Fresnel, Maxwell, Einstein) thought they were disagreeing about a basic constituent of our universe—what kind of entity makes up light?—but they were wrong. This view of scientific language seems to violate the weakest principles of interpretive charity because it makes scientists so deeply ignorant about the content of their own theories. It's not just that scientists can s-refer to things that they're ignorant about, which is a perfectly reasonable contention. It's that the possible world they would confidently identify as the one that contained their posits is not the possible world that contains their posits. Further, this view forces us to take some of the most interesting disputes in the history of science about what exists and radically reinterpret those disputes so that the participants are (unwittingly) positing the same entities but just disagreeing about what those entities are like. A theory of scientific language must fit not only with our metaphysics, but also with our history of science. And this account of scientific language leads to bad history.

The second objection to the Orwellian semantics embraced by the proponent of the metaphysical zoo is that it doesn't save a version of realism worth saving. The best the realist who embraces the metaphysical zoo can hope for is this: our theories are true but we probably don't know what our theories say. In other words, while our best theories

are true, we have no reason to think we understand our theories any better than Maxwell or Priestley understood theirs. One might call this view *We Don't Know What We're Talking About Realism*. The metaphysical zoo is a lot to swallow in order to save such a disappointingly vacuous form of realism.

4. Objections and clarifications

In presenting the direct version of the pessimistic induction to various audiences, I have found that it tends to elicit three strenuous objections. One objection, which itself tends to come in three forms, holds that reference is essential to the realism–antirealism debate, and so the direct pessimistic induction smuggles in reference as well. It merely avoids the *vocabulary* of reference. The second objection insists that reference is not essential to the realism–antirealism debate; and so the direct version of the pessimistic induction is merely a trivial variant on the extant versions of the argument. Therefore, it can add nothing to the debate. The third objection holds that the pessimistic induction can be defeated by a particular argument for scientific realism. I hope to clarify the direct pessimistic induction by briefly responding to each of these objections.

4.1. *Variations on the flight to reference*

I have offered a strategy for handling flight-to-reference arguments: insist on the distinction between s-reference and o-reference. While this might seem easy enough to do, it's really not. That's because there are many ways to sell a flight-to-reference argument so that it appears sensible and innocuous. I want to focus on three attempts to insist upon a flight to reference.

4.1.1. Straightforward ambiguity. Flight-to-reference realists might begin with an appeal to an intuitive connection between reference and existence: "Surely you will admit a connection between whether the central posits of a theory exist and whether the central terms of a theory refer." Then the realist launches into a defence of his favoured theory of reference, and *voila*, the flight to reference is off. The trick to responding to such an appeal is to keep o-reference and s-reference distinct. One should, of course, admit a connection between existence and o-reference. But in arguing that reference is essential to the debate, the realist cannot be permitted then to exchange o-reference for s-reference—which is what happens as soon as the realist puts forth a substantive theory of reference. The realist trades a notion of reference that brings with it a specific ontological commitment (o-reference) for a notion that brings with it a non-specific ontological commitment (s-reference).

  Perhaps the best way to defeat this equivocation is to keep in mind the dilemma implicit in Figure 1. In order for the realist's theory of reference to support realism, it must yield the result that expressions for obsolete posits refer. In that case, either:

A. The flight-to-reference realist is unwilling to apply schema (O) to claims of successful reference, in which case we cannot take seriously the appeal to the reference–ontology connection ("Surely you will admit a connection between the question of whether the central posits of a theory exist and whether the central terms of a theory refer"); or

B. The flight-to-reference realist is willing to apply schema (O) to such claims of successful reference, in which case the realist is committed to the metaphysical zoo.

It is only by equivocating that the flight-to-reference realist can have it both ways. He uses o-reference to insist upon a connection between existence and reference, and then he uses s-reference to avoid the metaphysical zoo while arguing that obsolete expressions refer.

4.1.2. Making ontological judgements. In Section 2.1, I offered a way to make ontological judgements that (allegedly) didn't presuppose a substantive theory of reference: assume our current best theories are true, draw up a comprehensive list of all the posits of the relevant theory, and determine whether the (purported) obsolete posit is a posit of our best current theories. In order to successfully complete this investigation, it is essential to *understand* the words used by the respective theories. The need for one to *understand* the respective theories is the impetus the realist might use to launch the flight to reference. In order to understand claims about the posit of an obsolete theory, one has to know the meaning of the expression used to designate the posit. For example, in order to understand claims about ether, one has to know the meaning of the expression "ether". But the flight-to-reference realist might insist that knowledge of meaning essentially involves knowledge of reference. So it follows that a theory of reference is essential to figuring out whether the ether exists.

Taken seriously, this argument leads to a clearly absurd conclusion. It is reasonable to assume that in order to understand an expression, one must know its referential properties. One must know how to properly apply the expression to the world. But the argument demands much more than this. The argument assumes that in order to understand an expression, one must know a *theory* of reference. And that is absurd. Such a requirement would make language impossible. For how would one come to understand a theory of reference in the first place? We can perfectly well understand words without knowing a general theory about how reference works. Speakers have been doing so for centuries.

There is a weaker version of this argument that has considerably more promise. Rather than insisting that knowing a theory of reference is essential to understanding all (potentially) referring expressions, the realist might insist that such knowledge is essential to understanding a small subset of such expressions. Let's turn to this argument.

4.1.3. Hard cases. There is no guarantee that schemas like (O) or (O′) will always yield definite answers about whether some posit exists. We might understand the expression "T" and know what the world is like, and yet not know whether T's exist. Consider one of Putnam's famous thought experiments (1975). Suppose we discovered that what we've been calling "cats" are really robots. Would we say cats exist, but they're just different from what we thought they were? Or would we say cats don't exist? Even though I understand the word "cat", I have no settled opinion about Putnam's example.[9]

Unlike the last two variations on the flight-to-reference argument, this one raises a genuine and difficult issue. Let's grant that there will be obsolete posits that are hard cases. For such a posit, it won't be clear whether it doesn't exist or whether it exists but is different from what it was thought to be. Does the existence of hard cases leave room for a flight-to-reference argument? It might seem so. The flight-to-reference realist might well argue as follows:

How are we to handle these hard cases? It is difficult to imagine resolving them any way other than by appealing to a theory of reference. Take Putnam's cats. The only way to figure out whether cats exist in Putnam's scenario is to figure out the referential properties of the expression "cats"—how that expression is properly applied to the world. And that's what a theory of reference does. So on this view, a theory of reference isn't essential to settle *all* cases (which would lead to absurdity), but it is essential to settle the *hard* cases.

The first thing to note about this argument is that it is an appeal to ignorance. Such arguments are notoriously fallacious. Merely because we can't imagine another way to resolve these cases doesn't mean that there isn't one. In fact, Stephen Stich has suggested a number of ways in which hard cases might be resolved that don't involve a flight to reference. For example, some hard cases might simply remain unresolved because no one is interested in whether the posit exists; other cases might be resolved via implicit previous agreements about what conditions would have to obtain in order for a posit to exist; still others might be decided by social and political factors internal to the relevant science, or by broader social and political factors (Stich, 1996, pp. 65–71). Whether or not Stich's proposals have merit, it is clear that the flight-to-reference realist needs some *argument* to show that hard cases about whether something exists *must* be resolved by appealing to a theory about language—rather than by appealing to the rich scientific practice and social context surrounding the historical trajectory of the posit in question. So even for hard cases, the direct pessimist need not accept the inevitability of flight-to-reference arguments.

The second thing to note about this flight-to-reference argument is that it too depends upon an equivocation between s-reference and o-reference. It is important to recognize that if we intend to use a theory of reference to resolve the hard cases—that is, to determine whether certain posits *exist*—we need a theory of o-reference. A theory of s-reference might yield the result that the expression for a hard case s-refers, but it doesn't follow that the posit exists. But the theory of reference offered by the realist is either a theory of s-reference, in which case it can't help with the hard cases anyway, or it's a theory of o-reference, in which case the flight-to-reference realist is, once again, committed to the metaphysical zoo. That's because in order to render plausible the contention that obsolete theories are approximately true, the realist's theory of reference must yield the result that the expressions for obsolete posits refer. (See Figure 1.)

There may be no point to commenting further on an argument that is guilty of two fallacies (appeal to ignorance and equivocation). But even if we put these objections aside, it is not obvious that the problem of hard cases can be used to undermine the direct pessimistic induction. That's because the direct pessimist is not committed to showing that *all* posits of successful, obsolete theories definitely don't exist. She needs to show only that a significant number of the central posits of such theories don't exist. So as long as there are a lot of cases of definitely non-existent posits of successful obsolete theories, the direct pessimist is in a position to make her case. Further, a realist who wants to offer a flight-to-reference argument on the basis of the existence of hard cases must admit that there are many easy cases. That's because flight-to-reference realists must be able to defend their preferred theories of reference. And they do this by showing how well their theories account for the easy cases. So ironically, the evidence the realist needs to begin his flight-to-reference argument (i.e. the easy cases) may be all the evidence the antirealist needs to show that the argument fails.

4.2. *The eliminability of reference*

Flight-to-reference realists respond to the direct pessimistic induction by insisting that reference is essential to the debate. Another response insists upon the exact opposite:

> You think that realists and antirealists have mistakenly and unwittingly turned their debate into a debate about language. But you have taken their talk of reference too seriously. To talk about central theoretical terms referring is just a way of approaching the issue of whether a theory's posits exist or not.

If the debate over the pessimistic induction has been employing the notion of o-reference all along, then this objection is correct—reference is eliminable. Talk of o-reference is a long-winded way of talking about existence. But if reference is eliminable from the debate over the pessimistic induction, it is a mystery why those most intimately involved in the debate have relied so heavily on it. While the proponent of the reference elimination objection may have no explanation for this mystery, the direct pessimist does. Flight-to-reference realists respond to the pessimistic induction by proposing a theory of s-reference, not o-reference. If viewed as a theory of o-reference (which *is* eliminable from the argument), flight-to-reference realists would land in the metaphysical zoo. So the flight-to-reference realist is actually defending a theory of s-reference, which is *not* eliminable from the flight-to-reference argument. If this explanation for why realists have offered flight-to-reference arguments is correct, then the eliminability objection rests on an equivocation. So the best response to the claim that reference is eliminable from the realist's argument is disappointingly crude: put up or shut up. Perhaps the realists' critique of the pessimistic induction depends for its success on a commitment to the existence of posits that most of us believe have long ago been discredited on empirical grounds. But I doubt it (see Section 3).

4.3. *The brute force response*

The brute force response involves attempting to defeat the direct pessimistic induction by *assuming* that the central posits of our current theories exist and that the central claims of our current theories are approximately true. This begs the question since the conclusion of the direct pessimistic induction is that the central posits of our current theories probably don't exist and that the central claims of our current theories are probably neither true nor approximately true. One might think that brute force responses are easy to spot. Not always. To see this, we need to back up a bit and consider the state of the debate over the pessimistic induction.

The pessimistic induction offers reasons for thinking that our best theories are false; further, it undermines the central argument for scientific realism. This argument, the optimistic abduction or the "no miracles" argument, contends that the best explanation for the predictive, explanatory and pragmatic success of our best scientific theories is that such theories are approximately true. The pessimistic induction directly attacks this contention by showing that there are many successful theories that are not approximately true. So the explanation for the success of those theories *can't* be that they are true. Because there must be some explanation for the success of those theories that does not involve their being approximately true, the realist is not entitled to assume that the best explanation for the success of our current best theories is that they are approximately true. An analogy might be useful here. Suppose I contend that the best explanation for the Brazilian national soccer team's success in the World Cup is their

high-quality goalkeepers. You respond that this can't be right. Brazil has often had great success in past World Cups with some very mediocre goalkeepers. If true, your response successfully undermines my Brazilian abduction. From the fact that Brazil is successful in a World Cup, I am not entitled to infer that their team has high-quality goalkeepers. This is so even if you are unable to offer an alternative explanation for Brazil's success in World Cups.

Against this backdrop, we can now consider what is probably the most sophisticated brute force response—the retention strategy. The retention strategy is employed with great care and skill by Kitcher (1993) and Psillos (1996, 1999). Putting all subtleties aside, it involves showing that those features of obsolete theories that were responsible for, or essential to, their success are retained in our best current theories. Notice that the realist's retention strategy only says that new theories retain the *successful* parts of old theories; it doesn't say that new theories retain the *true* (or approximately true) parts of old theories. The direct pessimist can embrace the retention strategy. How? She can grant that our best current theories *retain* the successful parts of obsolete theories and still claim that we have overwhelming reasons for thinking that all those theories are totally *false* (i.e. not even approximately true). Here is another way to see this point. The conclusion of the retention strategy is a conditional claim: if our current theories are approximately true, then the past ones must be approximately true too. But this conditional claim says nothing about whether any theory is actually (approximately) true. The direct pessimist can accept it, use the direct pessimistic induction to deny the consequent, and argue by modus tollens that our current theories are not approximately true.

There is much to admire in the retention strategy, especially in the skilful hands of Kitcher and Psillos. But as a reply to the pessimistic induction, it clearly begs the question.[10]

5. Concluding remarks

In the early to mid-20th century, English-speaking philosophy took a Linguistic Turn. Part of this Turn involved viewing all philosophical questions as questions about language. The philosophers involved in the contemporary realism–antirealism debate are typically not proponents of the Linguistic Turn and would in fact reject the notion that philosophical questions are nothing more than questions about language. It is therefore remarkable that the debate about the pessimistic induction ever got turned into a debate about language. After all, the questions about realism and antirealism that pique the interest and occasionally stir the passions of philosophers and non-philosophers alike are "Do most of the posits of our best theories exist?" and "Is the world roughly the way our best theories say it is?" These questions are ultimately about what the universe is really like. It is a testament to the power and the legacy of the Linguistic Turn in philosophy that this debate could have so naturally turned into a dispute about the nature of reference.

In this article, I have argued that the pessimistic induction is best understood as a sub-argument in a grand dilemma. Either realism is true or it's false. If it's false, then it's false. But if it's true, then (by the pessimistic induction) it's also false. The pessimistic induction is perfectly suited to being a *reductio* of realism. That's because the most powerful articulation of the pessimistic induction begins by assuming that realism is true. By assuming that the posits of our best current theories exist and what they say about the world is true, the antirealist argues that the posits of obsolete successful

theories do not exist and that the central claims of those theories are neither true nor approximately true. This, in turn, gives us inductive evidence that should be acceptable to the realist for thinking that the posits of our best current theories do not exist and what they say about the world is neither true nor approximately true. If the best the realist can do in response to the direct pessimistic induction is offer lots of claims about language, about s-reference, then the direct pessimist will win the day on the deep and important issues that separate the realist and antirealist—issues of truth and ontology.

Acknowledgements

Notes

1. The distinction between o-reference and s-reference is somewhat similar to the distinction drawn by Kripke (1977) between semantic reference and speaker's reference and the distinction drawn by Grice (1969) between conventional meaning and speaker's meaning. While the distinction drawn in this section doesn't map perfectly onto either of these distinctions, I will not explore their differences here.
2. One might wonder whether (O) permits expressions to designate non-homophonically designated entities. In other words, does it permit us to say that "Hesperus" refers to Phosphorus? If not, then it is a thoroughly defective schema. A close reading of (O) makes clear that it does not rule out the possibility of non-homophonic reference. (O) sets forth a necessary and sufficient condition for whether a posit exists and whether a term successfully refers. We can use (O) to tell us that "Hesperus" refers because Hesperus exists. But (O) is silent about whether Hesperus is identical to Phosphorus and about whether "Hesperus" refers to Phosphorus.
3. The example is Phillip Bricker's.
4. In an interesting article, Marc Lange (2002) makes a subtle and important point about pessimistic inductions generally. From the fact that most Fs have been failures in the past, it doesn't follow that probably most Fs are failures now. That's because it's possible for the following propositions to be true: (a) most Fs have been failures; (b) at any moment in time most Fs have not been failures. Lange's memorable example is that while most people who have been baseball managers have losing records, at most times, most employed baseball managers have winning records. How can this be? Turnover. Losing managers are unemployed more quickly than winning ones. Lange's insight suggests that the pessimist (direct or indirect) must argue that at most times before (say) 1900, most successful theories were false. While I accept Lange's point, I will ignore it here. My goal in this article is not to argue for antirealism. That would require a huge amount of detailed historical scholarship. Rather, my goal is to offer a framework for running the pessimistic induction that cannot be undermined by a flight-to-reference argument. Can the direct pessimist answer Lange's challenge by using the framework developed here to show that at most times in the past, most successful theories were false? As far as I'm concerned, this is an open historical question.
5. This example is somewhat complicated by the fact that present-day physicists sometimes use the expression "light waves" to refer to photons. It is important to see that this is irrelevant to the issue under consideration. We want to know whether the posit of an obsolete theory exists. Whether we're asking after ether or phlogiston or light waves, the relevant question is whether the entity, state, kind, etc., posited by an obsolete theory actually exists. *What our current theories say about the world* is relevant to figuring out whether the posits of these theories exist. But *which expressions our current theories use to describe the world* is not relevant to figuring out whether the posits of these theories exist. The point here is neither original nor deep. Obsolete wave theories of light posited something called "light waves". Whether that posit exists does not depend on it having been *called* "light waves" as opposed to "burritos". So the fact that current scientists still use the expression "light waves" cuts no ice.
6. Of course, there is some sense in which dephlogisticated air exists. Since there is no phlogiston, all air is dephlogisticated air (just as all space is Leprechaun-free). But this is not helpful to the realist, since the

realist wants to associate dephlogisticated air with oxygen, not with just any old air. After all, the realist contends that what Priestley discovered and what he called "dephlogisticated air" was oxygen.

7. Perfectly parallel considerations can be used to show that the particle theory of light was also false (i.e. its central posits did not exist and its central claims were neither true nor approximately true).

8. Two points about (A). First, the pessimist will need schemas for sentences containing predicates that have more than one place. Second, we must restrict the predicate "F". It can't be a predicate like "does not exist" or "is a fictional character". Such predicates will not occur in the context of the direct pessimistic induction. After all, we're applying this schema to the central theoretical and explanatory claims of successful obsolete theories. Such predicates do not typically occur in scientific theories.

9. In fact, given how under-described Putnam's cat example is, I don't think it's possible to make reasonable judgements about it. I would claim that a substantive theory of reference that yields a view about Putnam's cat example, as described, can't be correct. Arguments for this very brash claim are best left for another day.

10. One might argue that I have failed to recognize the subtlety of the retention strategy. The *unassuming* retention strategy does not assume that current theories are true; rather, it holds only that some elements of past theories that contributed to their successes have been retained in later theories. Therefore, we have good reason to believe that those posits of our own theories that contribute to their success will be similarly retained. (This point was made by an anonymous referee.) If this is the correct way to understand the retention strategy, however, it's hard to see how it is a reasonable response to the direct pessimistic induction. The direct pessimist has an argument to the conclusion that those elements of the obsolete theories that are allegedly retained in current theories include *lots of posits that don't exist*. The direct pessimist can gladly grant the realist's theory of s-reference as well as claims about the s-referential continuity between obsolete and current expressions. This doesn't block in any way the pessimist's conclusion: the terms for the obsolete posits didn't o-refer, and therefore, the terms for the current posits probably don't o-refer either. If the realist cannot attack the direct pessimist's argument about the *ontological* status of the posits of obsolete theories, then the direct pessimist has a surprising move to make: the direct pessimistic induction together with the unassuming strategy suggest that *our best theories retain the ontologically failed parts of old theories*. What more could the direct pessimist want from the realist, besides abject surrender? Without addressing the direct pessimist's basic *ontological* claims, all else (e.g. s-reference and retention) is bunting.

References

BISHOP, M. & STICH, S. (1998) The flight to reference, or how not to make progress in the philosophy of science, *Philosophy of Science*, 65, pp. 33–49.

BOYD, R. (1984) The current status of scientific realism, in: J. LEPLIN (Ed.) *Scientific Realism* (Berkeley, University of California Press), pp. 41–82.

DEVITT, M. (1981) *Designation* (New York, Columbia University Press).

EVANS, G. (1983) The causal theory of names, *Proceedings of the Aristotelian Society*, 47, pp. 187–208.

FIELD, H. (1994a) Deflationist views of meaning and content, *Mind*, 103, pp. 249–285.

FIELD, H. (1994b) Disquotational truth and factually defective discourse, *Philosophical Review*, 103, pp. 405–453.

GRICE, H.P. (1969) Utterer's meaning and intentions, *Philosophical Review*, 78, pp. 147–177.

HARDIN, C.L. & ROSENBERG, A. (1982) In defense of convergent realism, *Philosophy of Science*, 49, pp. 604–615.

HORWICH, P. (1990) *Truth* (Oxford, Blackwell).

KITCHER, P. (1978) Theories, theorists and theoretical change, *Philosophical Review*, 87, pp. 519–547.

KITCHER, P. (1993) *The Advancement of Science* (New York, Oxford University Press).

KRIPKE, S. (1972) *Naming and Necessity* (Cambridge, MA, Harvard University Press).

KRIPKE, S. (1977) Speaker's reference and semantic reference, in: P.A. FRENCH, T.E. UEHLING & H.K. WETTSTEIN (Eds) *Contemporary Perspectives in the Philosophy of Language* (Minneapolis, University of Minnesota Press), pp. 255–276.

LANGE, M. (2002) Baseball, pessimistic inductions and the turnover fallacy, *Analysis*, 62, pp. 281–285.

LAUDAN, L. (1984a) *Science and Values* (Berkeley, University of California Press).

LAUDAN, L. (1984b) Realism without the real, *Philosophy of Science*, 51, pp. 156–162.

LEPLIN, J. (1997) *A Novel Defense of Scientific Realism* (New York, Oxford University Press).

LEWIS, D. (1970) How to define theoretical terms, *Journal of Philosophy*, 67, pp. 17–25.

LEWIS, D. (1972) Psychophysical and theoretical identifications, *Australasian Journal of Philosophy*, 50, pp. 249–258.

NAGEL, T. (1997) *The Last Word* (New York, Oxford University Press).

PSILLOS, S. (1996) Scientific realism and the 'pessimistic induction', *Philosophy of Science*, 63 (Proceedings), pp. S306–S314.

PSILLOS, S. (1999) *Scientific Realism: How Science Tracks Truth* (London, Routledge).

PUTNAM, H. (1975) The meaning of 'meaning', in: H. PUTNAM, *Mind, Language and Reality: Philosophical Papers, Volume 2* (Cambridge, Cambridge University Press), pp. 215–271.

PUTNAM, H. (1978) *Meaning and the Moral Sciences* (London, Routledge & Kegan Paul).

RUSSELL, B. (1919) Descriptions, *Introduction to Mathematical Philosophy* (London, George Allen & Unwin), pp. 167–180.

SEARLE, J. (1958) Proper names, *Mind*, 67, pp. 166–173.

STICH, S. (1996) *Deconstructing the Mind* (New York, Oxford University Press).

Note on contributor

Michael A. Bishop is Associate Professor at Northern Illinois University. *Correspondence:* Department of Philosophy, Northern Illinois University, DeKalb, IL 60115, USA. E-mail: mbishop@niu.edu