

The Flight to Reference, or How *Not* to Make Progress in the Philosophy of Science*

Michael A. Bishop†‡

Department of Philosophy, Iowa State University

Stephen P. Stich

Department of Philosophy, Rutgers University

The flight to reference is a widely-used strategy for resolving philosophical issues. The three steps in a flight to reference argument are: (1) offer a substantive account of the reference relation, (2) argue that a particular expression refers (or does not refer), and (3) draw a philosophical conclusion about something other than reference, like truth or ontology. It is our contention that whenever the flight to reference strategy is invoked, there is a crucial step that is left undefended, and that without a defense of this step, the flight to reference is a fatally flawed strategy; it cannot succeed in resolving philosophical issues. In this paper we begin by setting out the flight to reference strategy and explaining what is wrong with arguments that invoke the strategy. We then illustrate the problem by considering arguments for and against eliminative materialism. In the final section we argue that much the same problem undermines Philip Kitcher's attempt to defend scientific realism.

*Received September 1996; revised May 1997.

†Send requests for reprints to S. Stich, Department of Philosophy, Rutgers University, 26 Nichol Ave., New Brunswick, NJ 08901-2882; e-mail: stich@ruccs.rutgers.edu

‡Earlier versions of the arguments developed in this paper were presented to the philosophy of science discussion group at Canterbury University in Christchurch, New Zealand, the Beijing Forum for Philosophy of Science at the Institute of Philosophy of the Chinese Academy of Social Sciences, the Philosophy Colloquium at the Graduate Center of the City University of New York, and at conferences at the University of Utah and Humboldt University in Berlin. We are grateful to all of these audiences for much valuable feedback. Special thanks are due to Philip Catton, Steve Downes, Hartry Field, Heimir Geirsson, Philip Kitcher and an anonymous referee for *Philosophy of Science*.

Philosophy of Science, 65 (March 1998) pp. 33-49. 0031-8248/98/6501-0002\$2.00
Copyright 1998 by the Philosophy of Science Association. All rights reserved.

There is a common strategy among contemporary philosophers for resolving issues in the philosophy of science. It involves making philosophical issues turn on questions of reference. We will call this strategy “the flight to reference.” The thesis we will defend in this paper can be stated very simply: Whenever the flight to reference strategy is invoked there is a crucial step that is left undefended, and without a defense of this step, the flight to reference is a fatally flawed strategy for resolving philosophical issues. Despite its importance, the undefended move in flight to reference arguments almost always goes unnoticed and, to the best of our knowledge, no one has made a serious attempt to show how the move might be justified.

Those who invoke the flight to reference always rely on one or another version of what we will call a *substantive* account of reference, an account that takes reference to be some sort of complex relationship between referring terms and entities or classes of entities in the world.¹ Their arguments can always be analyzed into three separate stages. In the first stage, they adopt (and sometimes explain and defend) their favored substantive account of reference; they say what specific relation or relations must obtain between a referring term and an entity or class of entities in order for the former to refer to the latter. In the second stage, they argue that on their account of reference the relation obtains between some term that is important for the debate at hand and some object or class of objects in the world. Or, alternatively, they can argue that the relation fails to obtain between the important term and any object or class in the world. At this stage, assuming all has gone well, the appropriate conclusion to draw is a conclusion about reference. But the philosophical debates in which these appeals to reference are embedded are not themselves debates about the reference of a term. Rather, they are debates about ontology or truth or some other matter. The third stage of the flight to reference strategy is an attempt to close this gap. The theorist uses the conclusion about reference drawn in stage 2 as a premise in an argument whose conclusion is explicitly about

1. Causal-historical theories of reference like those advocated by Kripke (1972), Putnam (1975) and Devitt (1981) are prime examples of the sort of theory of reference that we will classify as “substantive.” So too are description theories like those defended by Russell (1919), Searle (1958) and Lewis (1970, 1972), and hybrid theories such those developed by Evans (1983) and Kitcher (1978, 1993). Deflationary accounts of reference of the sort defended by Field (1986, 1994) and Horwich (1990) are the principle examples of theories that do not count as “substantive” on our use of the term. The argument to be developed in this paper does not need a hard and fast distinction between substantive and non-substantive accounts of reference. It requires only that those who employ the flight to reference strategy adopt accounts of reference that are obviously substantive.

truth or ontology or some other matter. But in order to do this the theorist relies (often tacitly) on one of a family of principles about reference. These principles all look to be obvious and trivial. Indeed, they are so obvious that some might think they are analytic or constitutive for reference. No relation could plausibly count as the reference relation unless it satisfied these principles.

It is at exactly this point that the flight to reference comes to grief. Presumably because the principle or principles invoked in stage 3 seem to be constitutive for reference, those who adopt the flight to reference strategy never try to establish that the reference relation adopted in stage 1 satisfies the principle. And this is the fatal gap in all flights to reference. For if the principle really is constitutive for reference, then theorists cannot legitimately claim that their favored substantive relation actually *is* the reference relation unless they give us some reason to suppose that their relation satisfies the principle. On the other hand, if the principle is not constitutive for reference, then before invoking it theorists must give us some reason for supposing that their relation satisfies the principle. Without an argument that the relation endorsed in stage 1 satisfies the principle invoked (or more typically, assumed) in stage 3, the flight to reference can tell us nothing about ontology or truth. And thus it cannot do the philosophical work that those who invoke the strategy want it to do.

It is our belief that this fatal defect in the flight to reference strategy undermines many influential arguments in the philosophy of science and elsewhere in philosophy. In this paper we will focus on arguments in two domains. The first, to be discussed in Section 1, is the debate about eliminative materialism, where the flight to reference is invoked by writers on both sides of the issue. The second, to be discussed in Section 2, is the debate over scientific realism. There our focus will be on Philip Kitcher's sophisticated attempt to use the flight to reference to defend a version of scientific realism.²

1. The Eliminative Materialism Debate: Reference and Ontology. Eliminative materialism is the view that intentional states, like beliefs and desires, do not exist. This is an ontological thesis—a thesis about whether instances of particular kinds of states inhabit our universe. Framed in this way, it is not clear why the truth of eliminativism should be related to theories about how words connect up to the world. To see how the eliminativism issue comes to depend on a semantic issue, consider the following argument for eliminativism:

2. In another paper (Bishop and Stich, in preparation) we examine the role that the flight to reference has played in recent debates about moral realism.

- (1) Folk psychology is an empirical theory and, like any empirical theory, it consists of various substantive theses. Beliefs and desires are among the theoretical states posited by folk psychology, and terms like ‘belief’ and ‘desire’ can be viewed as the central theoretical terms in this theory.
- (2) Folk psychology is “a false and radically misleading conception of the causes of human behavior and the nature of cognitive activity” (Churchland 1984, 43).

Both of these claims are very controversial, of course. But let us suppose they are true. It follows that beliefs and desires are posits of a false theory. How is the eliminativist going to get from there to the conclusion that beliefs and desires do not exist? Here is another claim that is often invoked in eliminativist arguments.

- (3) Theoretical terms are like definite descriptions. They refer to (or are satisfied by) those things that have (most of) the properties specified by the theory. Thus,
 - (3a) The central theoretical terms of false and radically misleading theories do not refer to anything.³

At this point, the first stage of the flight to reference is in place. In (3) a substantive theory is proposed about the relation that must obtain between terms and things in the world if the former are to refer to the latter. The second stage of the flight to reference strategy follows from (1), (2) and (3a), which together entail

- (4) ‘___ is a belief’ does not refer to anything.

So now we have a conclusion about reference. To get from there to a conclusion about the existence of beliefs, we need some principle linking reference and existence. And for this third stage of the flight to reference, the following principle looks like an ideal candidate:

- (5) (x) Fx iff ‘F___’ refers to x.

What (5) says is that something is an F if and only if ‘F___’ refers to it.⁴ And from (4) and (5) the eliminativist conclusion follows:

3. Many philosophers on both sides of the eliminativism debate have advocated one or another version of the description theory of reference for theoretical terms. See, for example, Lewis 1970, 1972; Churchland 1984, 56; McGinn 1991, 150; Stich 1983, Ch. 1. For a more detailed discussion of the role that description theories of reference have played in the eliminativism debate, see Stich 1996, §4.

4. (5) should be interpreted as a schema whose instances include:

(5-p) (x) x is a pig iff ‘___ is a pig’ refers to x.
and

(6) $\neg(\text{Ex})$ x is a belief. Or, less formally, beliefs do not exist.

But now what about (5)? What justification do we have for it? Well, it is hard to think of a more *obvious* claim about reference. Indeed, (5) is one of those apparently trivial principles that looks to many to be *analytic* or *constitutive* for reference. We confess that we have never really understood the notion of a constitutive principle, and for familiar reasons we are deeply skeptical about the notion of analyticity (see White 1950, Quine 1953, Harman 1967). But none of these concerns are relevant here. For we are more than happy to concede that, for whatever reason, (5) is certainly a principle that any account of the reference relation must respect. If a putative reference relation fails to satisfy (5), it couldn't really be the reference relation. And here we confront the crucial gap in the flight to reference strategy. For at the first stage, in (3), the argument offers a substantive account of the reference relation for theoretical terms. But is it the *right* account? Given what we have just conceded about (5), it *can't* be the right account unless it makes (5) true. Does it? Well, perhaps. But we have been offered no argument at all for this. And it is far from *obvious* that the (putative) reference relation sketched in (3) will make (5) come out true. Indeed, advocates of other substantive accounts of reference, which specify relations that are not extensionally equivalent, must think that when reference is understood as in (3), (5) is just plain false.

To see all this a bit more clearly, it will be useful to consider how an anti-eliminativist might concede the first two premises of the eliminativists' argument, but use the flight to reference strategy to come to exactly the opposite conclusion. There is no need to invoke a hypothetical anti-eliminativist here, since there is a very real and very acute one who has adopted just this approach. William Lycan accepts the view that folk psychology is an empirical theory and that terms like 'belief' and 'desire' are among the central theoretical terms of the theory. He also thinks it is entirely possible that folk psychology will turn out to be seriously mistaken because beliefs do not actually have most of the properties that folk psychology attributes to them. But none of this inclines Lycan to accept the eliminativists' ontological conclusion, since he rejects the account of reference urged in (3) and (3a), and adopts a much more "liberal" account.

(5-b) (x) x is a belief iff '___ is a belief' refers to x .

It might be a bit better to state (5) as follows, invoking satisfaction rather than reference:

(5-s) (x) Fx iff 'F___' is satisfied by x .

But since the difference is of no importance in our arguments, we will stick with (5). We propose to avoid fussing about technical matters like this whenever possible.

I am at pains to advocate a very liberal view. Unlike David Lewis (1972), and unlike Dennett (1978) and Stich (1982, 1983), I am entirely willing to give up fairly large chunks of our commonsensical or platitudinous theory of belief or of desire (or of almost anything else) and decide that we were just wrong about a lot of things, without drawing the inference that we are no longer talking about belief or desire. To put the matter crudely, I incline away from Lewis's Carnapian and/or Rylean cluster theory of reference of theoretical terms, and toward Putnam's (1975) causal-historical theory. As in Putnam's examples of 'water', 'tiger', and so on, I think the ordinary word 'belief' (qua theoretical term of folk psychology) points dimly toward a natural kind that we have not fully grasped and that only mature psychology will reveal. I expect that 'belief' will turn out to refer to some kind of information-bearing inner state of a sentient being . . . , but the kind of state it refers to may have only a few of the properties usually attributed to beliefs by common sense. (Lycan 1988, 31–32)

On the account that Lycan "is at pains to advocate" a theoretical term, T, refers to objects of kind O by virtue of a certain kind of complex causal-historical chain that connects uses of T to O (or to instances of O). The right kind of chain combines *groundings*, in which a term is introduced by a speaker who is in an appropriate causal relation with some instance of the kind, and *transmissions*, in which the term is conveyed from speaker to speaker. For our purposes, the details of the causal-historical account of reference are not essential (see Putnam 1975, Kripke 1972, Devitt 1981). What is important is that in "inclining toward" a Putnam-style theory of reference, Lycan has made the first move in the flight to reference; he has endorsed a substantive account of reference specifying an empirical relation that must obtain for terms of a certain kind to refer to things in the world.

One important fact about causal-historical accounts of reference is that they give no support to (3a). Quite to the contrary. On causal-historical accounts it is entirely possible for a theoretical term to refer even though it is embedded in a radically mistaken theory. This is enough to scuttle the eliminativist version of the flight to reference strategy, since without (3a) the eliminativist cannot get (4), and the argument grinds to a halt. But Lycan is not content to stop there, since he has his own version of the flight to reference. On his view it will probably turn out there is a "kind of information-bearing inner state" that stands in the appropriate causal-historical relation to our current uses of the ordinary word 'belief'. And if this is right, then (4) is not just unsupported, it is false; '___ is a belief' *does* refer to something. This is the second stage in Lycan's version of the flight to reference.

The third stage in Lycan's version is entirely parallel to the third stage in the eliminativists' version. In both cases, the third stage must rely on (5) or something similar. And in both cases the necessary steps seem so obvious that it is hardly necessary to state them. If *s* is an instance of the appropriate kind of information-bearing inner state, then '___ is a belief' refers to *s*. From this and (5) it follows that

(7) *s* is a belief.

And from this it follows that

(8) (Ex) *x* is a belief. Or, less formally, beliefs do exist.

So Lycan uses the flight to reference not merely to challenge the eliminativists' argument but also to argue that their conclusion is false.

The problem with Lycan's argument comes at exactly the same place as the problem with the eliminativists' argument. Both of them take (5) to be obvious, and so do we. No word-world relation would count as the reference relation if it did not satisfy (5). But Lycan has given us no more reason than the eliminativists did to suppose that if we take reference to be the relation he favors, it will make (5) true.⁵ Moreover, there is nothing obvious or trivial about the claim that Lycan's relation will make (5) true. Since the causal-historical relation is not extensionally equivalent to the (putative) reference relation invoked in the eliminativists' version of the flight to reference, they cannot *both* make (5) true, though of course they could both fail to make (5) true. Moreover, these are not the only two games in town. In attempting to give an account of reference, 20th century philosophers have proposed a variety of non-extensionally equivalent relationships between referring terms and things in the world. At most one of these can make (5) true, and there is no guarantee that any of them does.

We think the moral to be drawn here is quite clear. Philosophers who wish to invoke the flight to reference strategy must *defend* the claim that the substantive account of reference they endorse will make (5) (and similar obvious principles about reference) come out true. It is not an assumption anyone who accepts a substantive account of reference gets for free.⁶ However, it is hardly ever recognized that the

5. Actually, it is not the schema (5) but instances of (5), like (5p) and (5b) in fn. 4 that can be true or false. Our claims about the truth of (5) on various accounts of reference should be taken as shorthand for claims about the truth of all appropriate substitution instances of (5).

6. It might well be the case that proponents of deflationary accounts of reference *do* get (5) for free. But deflationary accounts are useless in flight to reference arguments, since they provide no way of arguing for the second stage without begging the question. To see why this is the case, a bit more detail about deflationary theories is needed.

assumption *needs* to be defended and, to the best of our knowledge, no one has ever tried. We do not claim to have an argument showing that it is impossible to defend assumptions of this sort. But since we have no idea how one would even begin to construct such an argument, and as far as we know no one else has ever tried, we are inclined to be more than a bit skeptical. What we do claim is that without a defense of this essential assumption, attempts to invoke the flight to reference are fatally flawed.

2. The Scientific Realism Debate: Reference and Truth. Scientific realism is defined by a cluster of theses, the most important of which is that successful, mature scientific theories are true or approximately true. Why should we believe this? The standard justification for scientific realism is an abductive argument: Mature scientific theories exhibit great explanatory and predictive success, and they facilitate our effective interventions in the world. The best explanation for this success is that such theories are *true* (or approximately true) representations of reality. Versions of this abductive argument can be found in many places including Smart 1968, Putnam 1978, Boyd 1984, and Kitcher 1993.

The “pessimistic induction” is perhaps the most serious challenge to

On deflationary accounts, predicates like ‘___ is true’ and ‘___ refers to ___’ exist “solely for the sake of a certain logical need” (Horwich 1990, 2). Reference “is merely a device for semantic ascent” (Horwich 1990, 8), it “is not a complex relation; a naturalistic or conceptual reduction is not needed and should not be expected” (Horwich 1990, 121). According to deflationists, a schema like (5) specifies what the word ‘refers’ means. If this is right, then deflationists who do not share our qualms about analyticity might well claim that (5) is analytic. This is half of what we had in mind when we conceded that deflationists might get (5) for free. The other half is that deflationists do not make any other systematic claims about reference. Since they do not think there is any “unified conceptual or naturalistic” (Horwich 1990, 124) reduction of reference, (5) captures all there is to be said about the reference relation in general. Thus there is no need to argue that the deflationists’ substantive account of reference makes (5) come out true. They do not offer a substantive account.

If (5) and its instances are all we have to work with, however, there is no way to bring off the second stage of a flight to reference argument. Consider, for example, Lycan’s flight to reference argument against eliminativism. He assumes (5-b):

(5-b) (x) x is a belief iff ‘___ is a belief’ refers to x.

and then uses the causal-historical account of reference to argue that there are states which satisfy the right hand side of the bi-conditional. No such move is available to a deflationist. The only way in which a deflationist can establish that ‘___ is a belief’ refers to a “kind of information-bearing inner state of sentient beings” is to go through the left hand side of (5-b) and argue that the kind of state in question is a belief. But, of course, that is exactly the claim that the flight to reference argument is supposed to establish. So deflationists cannot invoke the flight to reference without begging the question.

scientific realism. Larry Laudan (1984) presents this challenge in dramatic fashion by proposing a long list of theories that had great explanatory, predictive, and pragmatic successes but were *false*, and not just in detail. The laws and explanations proposed in these theories invoked substances, entities, and processes that do not exist. Among the examples Laudan cites are catastrophist geology, theories of spontaneous generation, the humoral theory of medicine, the effluvial theory of static electricity, the phlogiston theory of chemistry, the vibratory theory of heat, the vital force theories of physiology, and the theory of circular inertia (121). Laudan is very explicit about the lessons to draw from the history of science. “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” (121). “Since realists would presumably insist that many of the central terms of the theories enumerated above do not genuinely refer, it follows that none of those theories could be approximately true (recalling that the former is a necessary condition for the latter).” (122–123)

In *The Advancement of Science* (1993), Philip Kitcher adduces an ambitious defense of scientific realism. A central part of Kitcher’s response to the pessimistic induction is a version of the flight to reference strategy. Before we get to that, however, we will have to back up a bit and explain how Kitcher prepares the ground for his flight to reference. Kitcher begins by considering an example that looms large in Laudan’s work, 19th century wave theories of light. These theories held that light consists of waves propagated in a pervasive fluid-like medium, the aether. Laudan notes that “the optical aether functioned centrally in explanations of reflection, refraction, interference, double refraction, diffraction and polarization” as well as in “some very startling predictions” (e.g., Fresnel’s spot) “that, when tested, proved correct” (113–114). To prevent historical facts like the undeniable success of wave theories of light from undermining scientific realism, Kitcher’s first step is to propose a distinction between a theory’s *working posits*, which play an essential role in the theory’s success, and its *presuppositional posits*, which do not. According to Kitcher, light waves (also known as “aether waves,” “aether vibrations,” and “electromagnetic waves”) were a working posit of wave theories, while the aether was only a presuppositional posit.

Distinguish two kinds of posits introduced within scientific practice, *working posits* (the putative referents of terms that occur in problem-solving schemata) and *presuppositional posits* (those entities that apparently have to exist if the instances of the schemata

are to be true). The ether is a prime example of a presuppositional posit, rarely employed in explanation or prediction, never subjected to empirical measurement (until, late in the century A. A. Michelson devised his famous experiment to measure the velocity of the earth relative to the ether), yet seemingly required to exist if the claims about electromagnetic and light waves were to be true. The moral of Laudan's story is not that theoretical positing in general is untrustworthy, but that presuppositional posits are suspect. (Kitcher 1993, 149)

If Kitcher is right about all of this, then scientific realists need not worry about the fact that historically successful theories posited entities that do not exist, so long as the posits are presuppositional. But what about the *working* posits of theories that were once successful but that have now been abandoned? Kitcher is careful not to make any sweeping ontological claims; he does not say that the working posits of such theories always exist. What he does instead is focus on *reference*. To defend his version of scientific realism from the challenge posed by abandoned theories that were successful in their prime, Kitcher proposes to argue that many tokens or utterances of terms for working posits in such theories *succeed in referring*, and thus that many utterances of sentences containing these terms can be true, and often are. So, for example, Kitcher claims that sometimes a wave theorist's "tokens of 'light wave' . . . genuinely refer to electromagnetic waves of high frequency" (1993, 146), and on some occasions, "Priestley's token of 'dephlogisticated air' refers to . . . oxygen" (1993, 100). Often the statements that Priestley made when he used 'dephlogisticated air' were true, and indeed some of these statements expressed important new discoveries. What we find in the writings of the phlogistonians, Kitcher maintains, are "true doctrines trying to escape from flawed language" (1993, 100).

To support all of this, Kitcher (1978, 1993) sets out an elegant version of the flight to reference. In the first stage he advances a new, context sensitive, hybrid theory of reference in which the distinction between the reference of expression-types and the reference of expression-tokens plays a central role. In Kitcher's theory an expression-type like 'light wave' or 'dephlogisticated air' will have a number of different "modes of reference"—a number of different ways of connecting, or failing to connect, with nature. One token of 'light wave' with a given mode of reference might refer to electromagnetic radiation, while a different token with a different mode of reference might fail to refer to anything.

Modes of reference . . . fall into three types. A token's mode of reference is of the descriptive type when the speaker has a dominant

present intention to pick out something that satisfies a particular description and the referent of the token is whatever satisfies the description. The baptismal type is exemplified when the speaker has a dominant present intention to pick out a particular present object (or a set of objects, one member of which is present). Finally, the conformist type covers those (many) instances in which the speaker intends that her usage be parasitic on those of her fellows (or her own earlier self), and, in this case, the reference of her token is determined through a long causal chain that leads back to an initial usage, a usage in which a token produced by a first user has its reference fixed either in the descriptive or in the baptismal mode. (Kitcher 1993, 77–78)

In the second stage of his flight to reference, Kitcher applies his theory of reference to episodes in the history of science. He describes cases in which different utterances or tokens of obsolete terms refer on some occasions and fail to refer on others. For example, when Priestley used ‘dephlogisticated air’ with the intention to refer to “the substance obtained when the substance emitted in combustion is removed from the air” that token of the term was empty; it referred to nothing. Its mode of reference is of the descriptive type, and the description is not true of anything—there is no substance (no phlogiston) emitted during combustion that anyone could remove from the air. On another occasion, when Priestley used the term ‘dephlogisticated air’ with the dominant intention to refer to the substance that he and the mice he was using breathed in a specific experimental setting, the mode of reference of the token was of the baptismal type. And on those occasions “Priestley’s token of ‘dephlogisticated air’ refers to the substance which he and the mice breathed—namely oxygen” (1993, 100). In this example, Kitcher’s theory of reference, along with the facts about Priestley’s intentions, entails that a token of a term for a working posit succeeds in referring even though contemporary science regards the term as obsolete. This, of course, is just the sort of conclusion needed in the second stage of a flight to reference.

In the third stage of Kitcher’s flight to reference, the focus moves from reference to truth. Kitcher concludes, for example, that “Priestley enunciated various true statements which had not previously been accepted” (1993, 99). His general conclusion is that proponents of obsolete but successful theories often succeeded in discovering important new truths, and they also succeeded in stating those truths despite their use of obsolete, theory-laden terms like ‘aether waves’ and ‘dephlogisticated air’. If this is right, then an interesting form of scientific realism survives the pessimistic induction. Laudan’s picture of obsolete but

successful scientific theories—a picture on which most of their claims are false—is much too simple and very misleading. On Kitcher’s alternative picture, many tokens of obsolete theoretical terms succeeded in referring to things that really do exist, and many of the claims made by those who advocated these theories were true. But how, exactly, can Kitcher get from the stage 2 conclusions about reference to the stage 3 conclusions about truth? We very much doubt that he can do it at all.

As usual, our objection will focus on the tension between the first and third stages. For simplicity’s sake, let us use a slightly tidied up hypothetical example. On a certain occasion (we will call it “o”), Priestley used ‘dephlogisticated air’ with a baptismal mode of reference, intending to refer to the substance he had just produced in a particular experimental procedure. The first two stages of Kitcher’s flight to reference give us the following three premises:

- (9) On occasion o, Priestley uttered, “dephlogisticated air supports combustion better than ordinary air.”
- (10) On occasion o, “dephlogisticated air” refers to oxygen.
- (11) Oxygen supports combustion better than ordinary air.

All three of these premises are entirely unproblematic. (9) is assumed to be a historical fact; (10) follows from Kitcher’s theory of reference along with the assumptions we are making about occasion o; and (11) is a fact endorsed by contemporary chemistry. From these three premises, Kitcher wants to draw the following conclusion:

- (C) On occasion o, Priestley’s utterance, “Dephlogisticated air supports combustion better than ordinary air” is true.

But (C) does not follow from the above three premises, at least not directly. The inferential gap becomes evident when we note that the conclusion states that an utterance is true, but the premises say nothing about the conditions under which utterances are true. In order for this argument to succeed, then, we need some principle that connects reference and truth for utterances. An obvious candidate would be something like the following.

- (12) An utterance of the form ‘Fa’ is true iff (Ex) (this token of ‘a’ refers to x and x satisfies this token of ‘F___’).

Actually, this is not quite enough to get Kitcher’s conclusion. To make all the pieces fit together, we need to replace (11) with

- (11’) Oxygen satisfies ‘___ supports combustion better than ordinary air’.

And one might ask how, exactly, we can justify the replacement.⁷ This is not a problem we propose to press, however, since as we see it the real problem with the argument is centered on (12). What justification do we have for (12)? The most natural answer, we think, is that (12) hardly needs justification. Indeed, it is yet another of those claims about reference that some theorists might take to be analytic. And while we are already on record as being skeptical about analyticity, we would happily concede that (12) is *obvious* in much the same way that (5) was. No account of the reference relation that failed to make (12) true could possibly be correct.

So what is the problem? The answer, of course, is entirely parallel to the answer we gave in our critique of the eliminativist and anti-eliminativist versions of the flight to reference. Kitcher has given us a substantive theory of reference. It is a subtle and complex account that differs in various ways from the accounts offered by other theorists. What he has not done, indeed what he has not even begun to do, is argue that the complex relation that he calls “reference” makes all instances of (12) come out true. And it is certainly not *obvious* that Kitcher’s relation does this. For as noted earlier, there are lots of substantive accounts of reference to be found in the literature, and most of them specify relations that are not extensionally equivalent to the one Kitcher so carefully describes. Moreover, the relation specified by some of these competing accounts (particularly those that emphasize the “descriptive mode”) will make the right hand side of some instances of (12) come out false when Kitcher’s account makes it come out true, while the relation specified by other competing accounts (particularly those that emphasize the “baptismal mode”) will make the right hand side of some instances of (12) come out true when Kitcher’s account makes it come out false. Let $R_{\text{descriptive}}$ and $R_{\text{baptismal}}$ be two such alternative relations. Then if Kitcher’s relation makes all instances of (12) come out true, neither of the others do, while if either $R_{\text{descriptive}}$ or $R_{\text{baptismal}}$ makes all instances of (12) come out true, then Kitcher’s relation does not. Without some *argument* for the claim that *his* relation makes all instances of (12) come out true (and thus that the others do not) Kitcher has no justification for invoking (12). And without (12) or something like it, Kitcher cannot get from (9), (10) and (11) to (C); he cannot get from premises about the complex relation that he calls “reference” to conclusions about truth. The most important element

7. Alternatively, we could use a less general version of (12) like

(12') An utterance of the form ‘a supports combustion better than ordinary air’ is true iff (Ex) (this token of ‘a’ refers to x and x supports combustion better than ordinary air.)

of scientific realism is the claim that successful scientific theories are true (or approximately true), and Kitcher's elaborate flight to reference argument provides no justification at all for this claim.⁸

A defender of Kitcher's flight to reference arguments might protest that similar argumentative strategies are common in science.⁹ Consider the classic example of Neptune's discovery. In the early 1800s, astronomers found that the observed orbit of Uranus did not fit the best available predictions. By 1836 "most astronomers had accepted the hypothesis of an exterior planet" to account for the discrepant observations (Grosser 1979, 54). About seven years later, John Couch Adams and Urbain Jean Joseph Leverrier independently calculated approximately where the new planet would be found in the sky. Although Adams assured English astronomers that "the planet would appear no smaller than a star of the ninth magnitude" they "elected to map all stars down to the eleventh magnitude" (Grosser 1979, 108). Across the Channel, Leverrier assured French astronomers that they would not have to map all the stars since the disk of the new planet would be very distinctive. But no French observatories searched. In frustration, Leverrier wrote to an assistant at the Berlin Observatory, J.G. Galle, asking him to look for Neptune. The letter reached Galle on 23 September 1846. That evening, Galle described what looked to be a star of the eighth magnitude that was not listed in his star charts. He had found Neptune.

Here we have an instance of scientific reasoning from some phenomena (the observed discrepancy in Uranus' orbit) to a theoretical posit (the planet Neptune) and then to a further conclusion (Neptune's apparent magnitude and location). Kitcher's flight to reference seems to employ just this kind of argumentative strategy: from some phenomena (various facts about a scientist's use of a particular expression on a particular occasion) to a theoretical posit (a certain expression successfully refers) and then to a further conclusion (a scientist's utterance

8. It might be thought that Kitcher's argument could sustain a somewhat less central element of scientific realism, namely the claim that terms for the working posits of successful theories succeed in referring—or at least that some tokens of such terms succeed in referring. But even this is problematic. For if it is indeed the case that no account of the reference relation that failed to make (12) true could possibly be correct, then without an argument that *his* account of reference makes (12) true, Kitcher has no grounds for claiming that the complex substantive relation he describes really is reference. Though he has established that certain tokens of 'dephlogisticated air' stand in the relation he describes to oxygen, he is not entitled to draw any conclusions about the *reference* of those tokens until he offers an argument that his relation makes (12) and other obvious (or "constitutive") principles about reference true.

9. The line of defense sketched here was suggested to us by Philip Kitcher.

on a particular occasion is true). If the former strategy is acceptable (and we certainly agree that it is), then isn't the latter strategy also acceptable? The answer, we maintain, is that it is not because there is a crucial disanalogy between the strategies. In the scientific case, *both inferences were based on extraordinarily well-confirmed empirical theories*. A well-placed confidence in various theories, including Newtonian physics, as well as many careful observations led scientists to conjecture that Neptune exists. And on this assumption, Adams and Leverrier could reasonably infer many things about the undiscovered planet—its mass, its distance from the Earth and Sun, its apparent magnitude and location—once again, all on the basis of extraordinarily well-confirmed theories.

If we take this analogy seriously, we should expect each inferential step in the flight to reference strategy to be supported by well-confirmed empirical theories. But they are not. Indeed, one way to see the point of our argument more clearly is to note how the flight to reference strategy utterly fails to meet the epistemic standards set by Adams and Leverrier. In Kitcher's flight to reference, principle (12) supported the second inference, from reference to truth. What justifies it? Kitcher, like all other proponents of flight to reference arguments, does not offer (12) as a principle that is part of, or implied by, a well-confirmed empirical theory. We suspect that many philosophers assume that principles like (12) are constitutive for reference, which would explain why no one has proposed an empirical theory that would justify these principles. For argument's sake, we are prepared to grant this assumption. But now, those who urge arguments like Kitcher's face a dilemma:

- (i) If Kitcher claims that (12) is constitutive for reference, he gets the second inference for free, but he still needs to justify the first one. What reason is there to believe that the word-world relation specified (or defined?) by Kitcher's theory actually satisfies principle (12), which is *by hypothesis* constitutive for reference? As is typical in flight to reference arguments, no well-confirmed empirical theory is in sight. But an argument of some sort is essential. For it cannot simply be *stipulated* that the word-world relation Kitcher specifies has the property of making (12) true, any more than Adams could simply stipulate that Neptune would appear no smaller than a star of the ninth magnitude.
- (ii) If Kitcher does not claim that (12) is constitutive for reference (or that (12) is obvious and that any acceptable account of reference must make it come out true), then we will offer no objection to the first inferential step. But what about the sec-

ond inference? What reason is there to believe that the word-world relation specified by Kitcher's theory actually satisfies principle (12)? After all, *by hypothesis* (12) is not constitutive for reference. Once again, some sort of argument is essential.

If we take the proposed analogy with scientific reasoning seriously, those who invoke the flight to reference owe us full-blown, well-confirmed empirical theories that support the contention that their favored reference-relations satisfy principles like (12). But we are willing to settle for less. Any sort of cogent argument will do.

3. Conclusion. Logical positivists believed that all philosophical questions are really questions about language. While the tenets that lead to this view of philosophy have faded, a remnant of the view has survived and flourished: the temptation to make philosophical issues turn on questions of reference. We have described one common strategy for trying to resolve philosophical issues by appeal to substantive accounts of reference, the strategy we have called the flight to reference. And we have argued that there is a fatal gap in all arguments employing this strategy. In order for a claim about reference to be relevant to claims about existence or truth or some other philosophically important notion, a principle linking reference to that notion is required. However, those who rely on such principles face a dilemma. Either they take the principles to be obvious constraints on any substantive account of reference, or they do not. If they do, then they are simply not entitled to claim that their own favored substantive relation really is the reference relation unless they give us some reason to suppose that their relation will make the relevant principle true. If they do not, then they are free to claim that their favored relation is indeed reference, but they cannot invoke the principle until they argue that when reference is identified with the relation they have specified, it makes the principle come out true. Either way, then, a theorist who invokes the flight to reference must argue that when the substantive reference relation endorsed in the first stage of the flight is plugged into the principle relied upon in the third stage, the relation makes the principle come out true. And that, we suspect, will be no easy task.

REFERENCES

- Bishop, M. and S. Stich (in preparation), "Moral Realism and the Flight to Reference".
 Boyd, R. (1984), "The Current Status of Scientific Realism", in Jarrett Leplin (ed.), *Scientific Realism*. Berkeley: University of California Press, pp. 41–82.
 Churchland, P. M. (1984), *Matter and Consciousness*. Cambridge, MA: MIT Press.
 Dennett, D. (1978), *Brainstorms*. Cambridge, MA: MIT Press.
 Devitt, M. (1981), *Designation*. New York: Columbia University Press.

- Evans, G. (1983), "The causal theory of names", *Proceedings of the Aristotelian Society* 47: 187–208.
- Field, H. (1986), "The Deflationary Concept of Truth" in G. MacDonald and C. Wright (eds.), *Fact, Science, and Value*. Oxford: Blackwell, pp. 55–117.
- . (1994), "Deflationist Views of Meaning and Content", *Mind* 103: 249–285.
- Grosser, M. (1962/1979). *The Discovery of Neptune*. New York: Dover Publications.
- Harman, G. (1967), "Quine on Meaning and Existence I", *Review of Metaphysics* 21: 125–151.
- Horwich, P. (1990), *Truth*. Oxford: Blackwell.
- Kitcher, P. (1978), "Theories, Theorists and Theoretical Change", *Philosophical Review* 87: 519–547.
- . (1993), *The Advancement of Science*. Oxford: Oxford University Press.
- Kripke, S. (1972), *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Laudan, L. (1984), *Science and Values*. Berkeley: University of California Press.
- Lewis, D. (1970), "How to Define Theoretical Terms", *Journal of Philosophy* 67: 17–25.
- . (1972), "Psychophysical and Theoretical Identifications", *Australasian Journal of Philosophy* 50: 249–258.
- Lycan, W. (1988), *Judgement and Justification*. Cambridge: Cambridge University Press.
- McGinn, C. (1991), "Mental States, Natural Kinds and Psychophysical Laws", in *The Problem of Consciousness*. Oxford: Basil Blackwell, pp. 126–152.
- Putnam, H. (1975), "The meaning of 'Meaning' ", in *Mind, Language and Reality*. Cambridge: Cambridge University Press, pp. 215–271.
- . (1978), *Meaning and the Moral Sciences*. London: Routledge & Kegan Paul.
- Quine, W.V. (1951), "Two Dogmas of Empiricism", *Philosophical Review* 60: 20–43.
- Russell, B. (1919), "Descriptions", in *Introduction to Mathematical Philosophy*. London: George Allen & Unwin Ltd., pp. 167–180.
- Searle, J. (1958), "Proper Names", *Mind* 67: 166–173.
- Stich, S. (1982), "On the Ascription of Content", in A. Woodfield (ed.), *Thought and Object*. Oxford: Oxford University Press, pp. 153–206.
- . (1983), *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press.
- . (1996), "Deconstructing the Mind", in *Deconstructing the Mind*. Oxford: Oxford University Press, pp. 3–90.
- Smart, J.J.C. (1968), *Between Science and Philosophy*. New York: Random House.
- White, M. (1950), "The Analytic and the Synthetic: An Untenable Dualism" in S. Hook (ed.), *John Dewey: Philosopher of Science and Freedom*. New York: Dial Press, pp. 316–330.