

EPSA'07

Section 1. General Philosophy and Methodology of Science

THEORIES OF REFERENCE AND THE PHILOSOPHY OF SCIENCE

Panu Raatikainen

Department of Philosophy, P.O. Box 9,

FIN-00014 University of Helsinki, Finland.

panu.raatikainen@helsinki.fi

The theses of meaning-variance and incommensurability of Kuhn and Feyerabend have consequences which are not easy to digest: that it does not make sense to say that a later scientific theory is more adequate, or is closer to the truth, than an earlier one, and that there is therefore no genuine progress in science. Now it is widely agreed that there must be something seriously wrong with this view, but there is no similar general agreement on what would be the best cure.

The radical ideas of Kuhn and Feyerabend derive from their assumption of the contextual theory of meaning. It is a version of the description theory or reference (or descriptivism, in short), according to which the descriptions that speakers analytically associate with an expression give the meaning of it, and determine its reference. As a consequence, Kuhn and Feyerabend rejected the very possibility of an independently meaningful observational language, a common ground from which the rival theories could be compared, as well as the possibility of crucial experiments. Even the meanings of alleged observational terms are, according to them, exhausted by their theoretical context.

It has sometimes been suggested that the so-called new theory of reference, or the causal theory of reference,¹ developed especially by Kripke (as a novel alternative to the more traditional description theory of reference), would provide an alternative picture of meaning and reference which avoids the unwelcome consequences of the meaning-variance thesis. Roughly, the idea of the causal theory of reference is that an expression refers to whatever is appropriately causally linked to the speaker. By cutting off the link between reference of a term and the beliefs (or, the theory) associated with the term, it allows the reference to stay stable even if the associated beliefs change. And the

¹ A terminological note: I shall use “the causal theory” for the specific picture which was put forward by Kripke as the correct account of the reference of most ordinary proper names, and perhaps of some general terms (see below), and I shall call the more general picture of reference that emerges from the writings of Kripke, Putnam, Schwartz, Devitt and others “the new theory of reference”. On this terminology, “the causal theory of reference” is a proper part of “the new theory of reference” (as becomes clearer below).

sameness of reference is sufficient for comparison between theories, even if the meaning changes.

This proposal has been welcomed by some (e.g. Boyd 1973; Kitcher 1978; Newton-Smith 1981; Hacking 1983; Devitt 1979, 1984/1999), but numerous philosophers of science have been quite critical towards the causal theory of reference (e.g. Fine 1975; Enc 1976; Mellor 1977; Papineau 1979; Nola 1980; Dupre 1981; Kroon 1987; Psillos 1999; Niiniluoto 1999; Bird 1998, 2000). Some favour instead “causal descriptivism” (e.g. Bird, Psillos).

Objections to the New Theory of Reference

Let us next look at what are the most influential objections to the new theory of reference, with a particular focus on the philosophy of science. To begin with, the standard critical claim – especially among the philosophers of science – is that the new theory of reference fails to account for reference failure. That is, it is proposed that the new theory of reference cannot explain the cases, not absent in the history of science, where one has ended up with the conclusion that certain postulated theoretical entities did not exist after all. But certainly the scientists who first introduced the term were causally interacting with something that caused the phenomenon which the entities were postulated to explain. For example, it is now agreed that “phlogiston” failed to refer to anything real. But there was something, namely oxygen, present in combustion. Therefore, the argument continues, if the new theory of reference were true, “phlogiston” should refer to oxygen rather than fail to refer (e.g. Enc 1976; Nola 1980; Kroon 1985; Niiniluoto 1999, p. 126; Psillos 2000, p. 290; Bird 2000, p. 185).

Further, a usual objection for the new theory of reference, in the context of the philosophy of science, is that ostension, or perceptual contact, important in the introduction of a name as the new theory of reference describes it, is just not possible in the case of theoretical entities of science – in other words, that theoretical entities are not easily pointed to. Hence, it is concluded, the new theory of reference cannot explain the reference of theoretical terms (see e.g. Bird 2000, p. 184)

Finally, Papineau (1979), Dupre (1981), and many others, have complained that a sample will usually be a member of many kinds. So how can a *general term* be introduced? If it happens via the initial baptism in the ostensive contact with a sample, as the causal theory of reference seems to suggest, how can one rule out wrong kinds of generalizations? This is the so-called *qua*-problem (see Devitt & Sterelny 1999, 72-75). For example, a particular tiger is not only a sample of a tiger, but in addition, say, of a Siberian tiger, as well as of a feline, a mammal etc.

These seem to be the main objections to the new theory of reference among the philosophers of science. I shall argue that many of them have an over-simplified and, in part, mistaken understanding of what the new theory of reference, or the causal theory in particular, amounts to. I shall briefly review the principal ideas of the causal theory, and explain how the new theory of reference can account for reference failure, or for reference to unobservable theoretical entities. I also argue that causal descriptivism is in any case a non-starter.

A Closer Look at the New Theory of Reference

In order to evaluate how well the above critiques hit their target, it is necessary to take a bit closer look at the causal theory of reference (the positive account is mainly due to Kripke). The theory has two parts: a theory of introduction of referring expressions, and a theory of subsequent reference transmission, or, 'reference-borrowing' (for simplicity, let us first consider only proper names):

First, there is the initial introduction of a referring expression to the language, a baptism or a dubbing event. There, an object must somehow be singled out for naming. According to Kripke, this can happen either with the help of a description, or of an ostension (ie., by pointing to it). Moreover, Kripke even adds: "The case of baptism by ostension can perhaps be subsumed under the description concept also. Thus the primary applicability of the description theory is that of initial baptism." (Kripke 1980, p. 96, fn 42). It must be added, though, that in the case one is in direct perceptual contact with the object named, an additional description used may not be satisfied by the referent.

Second, other speakers not present at the name-giving occasion acquire the word from those in attendance at the baptism, still others from the former, and so on. Later users of the name need not know or be able to identify the referent. They acquire the name from earlier users of the name, and it is sufficient for successfully referring to the referent that they are part of an "historical" or "causal" chain of speakers which goes back to the first users. Speakers may be largely ignorant of this chain, and even from whom they got the name. Nevertheless, they can successfully refer using the expression. This is the phenomenon of reference borrowing.

In other words, any description used in the *short-term* to pick out the object to be named need not retain any *long-term* association with the name. So long as the intention remains to keep on referring to the same individual to which one has been referring, the description may be forgotten or misremembered. But Kripke's *main* point is that no uniquely identifying description need be or generally is transmitted along with a name (cf. Burgess 2006).

Interim Conclusions

Two important points deserve emphasizing here: First, it is of crucial importance to note that the causal relation, if it is mentioned at all (for many of those advocating the new theory of reference rather prefer to talk about an “historical chain”), is said to hold between earlier and later *users* of the name (and even then, Kripke often puts “causal” inside quote marks). The thesis is emphatically not that there has to be a causal connection between the object referred to and a speaker, much less that an expression refers to whatever is causally responsible for the uses of the expression, or to whatever with which the introducers of the expression were causally interacting when it was first introduced.

Therefore, when philosophers of science make claims such as “the causal theory was supposed to identify the reference of a term with the cause of our use of the term” (Bird 2000, p. 185), or that “the causal theorists insist that what fixes the reference of a term ... is the causal chain which connects the term with the *object* named in the dubbing ceremony” (Psillos 1999, p. 282), they are simply misrepresenting the essential idea of the causal theory of reference. Thus, it begins seriously to look like the “causal theory of reference” of many philosophers of science is in actuality a straw man, a caricature of the real theory. Consequently, it is also not as clear as many have thought that it is in trouble with reference failure (see also below).

Second, in the initial introduction of a term, it is clearly allowed by Kripke, and all those who follow him, that a description is used to single out the referent. If the description then fails, one may fail to name anything from the outset, and the resulting name is empty. For Kripke, anything that can be described can be named. The description need not involve ostension (though it may). Thus, according to Kripke, even numbers and other abstract entities can well be named, with the help of descriptions. Neither perceptual contact nor any causal connection is necessarily required, in order for one to be able to introduce a referring name. To repeat, the “causal” chain then occurs between the first and later users of the name.

The Scope of the Causal Theory

It is also important to recognize that the causal account was never proposed as a universal theory of referring expressions without exceptions, but primarily as a theory of ordinary proper names and *observational*, or manifest, natural kind terms such as “water”, “gold”, or “tiger”.² It has been allowed from the beginning that some expressions, even some

² Devitt (1981), for example, explicitly talks about “*observational* natural kinds” here. More recently, many philosophers of language have started to call such kinds “*manifest* natural kinds” (e.g. Soames). I shall use, in the present context, “observational”, in accordance with the terminology in the philosophy of science.

proper names, may refer according to the lines that the description theory suggests. Thus, Kripke suggested that, for example, “Jack the Ripper” may be taken to refer to the man, whoever committed all these murders, or most of them (Kripke 1980, p. 79). He only adds: “But in many or most cases, I think the thesis [descriptivism] is false” (*ibid*, p. 80)

Putnam in turn allows that there are – in addition to observational natural kind terms, which (with proper names) function along the lines that the causal theory suggests – also law-cluster terms³ and even what Putnam calls one-criterion terms; the meaning of the latter can indeed be given by a simple definition; for example “vixen” means the same as “female fox”. Such words are, though, quite rare according to Putnam.

Devitt (1981) in turn introduces, generalizing Donnellan’s distinction, a systematic terminology for this same issue and talks about attributive and designational names. In this terminology, most ordinary names, like “Aristotle”, are designational (or referential) – but “Jack the Ripper”, for example, is attributive. The same distinction obviously applies to general terms.

More on Reference Failure and Theoretical Terms

Let us now reconsider the issue of empty names, or reference failure – the cases such as “phlogiston”. To begin with, it was already pointed out above that the causal account of reference has been put forward first and foremost as a theory of ordinary proper names and observational natural kind terms. However, “phlogiston” for example simply is not such an *observational* natural kind term; it was not introduced in a perceptual contact with a sample of the substance. An advocate of the new theory of reference may well submit that it is rather an attributive term, perhaps even comes close to what Putnam calls one-criterion term, the meaning of which is given (roughly) by the definition “the substance emitted in combustion”.

We may compare “phlogiston” (and such) to Kripke’s example of “Jack the Ripper”, which, according to even Kripke, can be taken to mean, roughly, the same as the description, “the man, whoever committed all these murders, or most of them” (Kripke 1980, p. 79). If it, however, turned out that each one of those murders was in fact committed by a different person or that there were no murders at all, but that these were all improbable accidents, one could conclude that ‘Jack the Ripper’ did not exist. Similarly, a proponent of the new theory of reference need not assume that phlogiston exists after all; the new theory of reference can well be accommodated to reference failures.

³ Though Putnam later changed his mind and concluded the the law-cluster picture gives a wrong account of observational natural kind terms, he never gave up the idea itself, and has continued to hold that this would be an appropriate account of some scientific terms. (see e.g. Putnam 1986, 1988)

Now what about the introduction of theoretical terms in general? We have already seen that the new theory of reference just does not demand that the introduction of a name must involve ostension, or perception – a pure description is sufficient. Hence, there is no real problem in introducing a theoretical term, no more than there was with introducing “Jack the Ripper”, from the point of view of the new theory of reference.

Causal Descriptivism

As was mentioned above, some philosophers of science prefer a new version of the description theory of reference, known as “causal descriptivism”, to the new theory of reference. This is typically motivated by the qua-problem. That is, it has been repeatedly suggested that in order to avoid the wrong sort of generalizations, a pure causal theory needs supplementation by a descriptive element, and that in turn leads us to causal descriptivism (see e.g. Sankey 1994; Psillos 1999; Bird 2000). Thus, one can read, for example: “On account of these difficulties the view has become widespread that the causal theory needs supplementation by a descriptive element, what we may call ‘causal descriptivism’” (Bird 2000, p. 185). However, apparently several distinct issues get conflated here.

To begin with, there is the possible descriptive element in the first introduction of a term. Now it has always been allowed by the advocates of the new theory of reference that there are descriptive elements in the determination of a name’s reference. The fundamental point of the new theory of reference is that such a descriptive element associated with a name is often insufficient alone for uniquely identifying the referent, or extension. This much was granted very clearly for example already in the standard textbook exposition of the new theory of reference by Devitt and Sterelny (1987), and as at least Sankey openly grants, “[a]rguably, this [descriptive elements] is already a feature of Putnam’s original account” (Sankey 1994, p. 71) (Moreover, although Kripke did not explicitly consider exactly this issue, there is no reason to think that he would have objected; after all, he even allowed the *introduction* of a term to be purely descriptive.) In particular, it has been long admitted that – in order to rule out wrong kinds of generalizations, i.e., the qua-problem – there must be descriptive elements in use in the introduction of an expression.

Thus Devitt and Sterelny concede: “the introducer of a name must use some general categorical term such as ‘animal’ or ‘material object’ ((Devitt & Sterelny 1987, p. 65); and: “It seems that *our* causal theory of names cannot be a ‘pure-causal theory’. It must be a ‘descriptive-causal’ theory” (Devitt & Sterelny, *ibid.*) (note, however, that such a ‘pure-causal theory’ has in reality been nobody’s theory; that is, no key advocate of the new theory of reference has ever held such a view). Sankey (1994) agrees with Devitt and Sterelny, but unfortunately calls the resulting view quite misleadingly “a form of causal

descriptivism". This may have caused confusion, for what one usually means by "causal descriptivism" is something quite different. Psillos and Bird, for example, seem rather to assume – at least in some points – that it is the standard interpretation of Causal Descriptivism, deriving from Lewis (1984) and Kroon (1987), which is assumed here.

In any case, allowing such a descriptive element does not amount to a return to the description theory of reference – not by a far cry. Accordingly, Devitt and Sterelny write: "Clearly, we have moved some distance back toward the description theories rejected earlier... However, the extent of the move should not be exaggerated. First, the association of a general categorical term certainly does not amount to identifying knowledge of the object. Second, our movement is a modification of the causal theory of grounding [i.e. name introduction]. The causal theory of reference borrowing remains unchanged; borrowers do not have to associate the correct categorical term" (Devitt & Sterelny 1987, p. 65).

As already noted, the descriptive element (a categorical term) at issue here is certainly not what "causal descriptivism" suggests. So what really then is causal descriptivism? One may express the basic idea of causal descriptivism, as it is usually understood, schematically, as follows:

Speakers associate a name "N" with a description of the form

(CD) The entity standing in relation R to my current use of the name "N",

and this description determines the reference of "N". The relation R here is drawn from the rival non-descriptivist (e.g. causal) theory of reference.

One may now note that if the causal theory of reference really were refuted by the alleged problems mentioned above, the same problems would rebut causal descriptivism as well. They stand or fall together. For example, if the empty names like "phlogiston" really were a devastating counter-example for the causal theory of reference, they would be just as much that for causal descriptivism, which builds the causal theory of reference in its descriptions. Similarly, causal descriptivism does not itself help in any way with the qua-problem. Further, causal descriptivism has various problems of its own (see e.g. Devitt & Sterelny 1999, p. 61; Raatikainen 2006).

There is a third theme which also seems to loom behind these debates and sometimes gets conflated with either the causal theory or causal descriptivism (and sometimes with the possible descriptive elements in name introduction). Namely, the causal theory of reference, or causal descriptivism (which mimics it), should not be confused with Putnam's idea of "causal descriptions" (e.g. Putnam 1973, p. 200). That is, Putnam has suggested that some terms, perhaps "electricity", for example, are first introduced with causal descriptions such as: "By *x* I mean the magnitude that is responsible for such and

such effects” or “By y let us mean the objects which are responsible for these and these observable phenomena” etc. In passing, Kripke made a somewhat similar suggestion, that perhaps “Neptune”, for example, was first introduced with the help of a causal description like “the planet which causes such and such discrepancies in the orbits of such and such other planets” (see Kripke 1980, p. 79, fn 33).

Such causal descriptions were, for Kripke and Putnam, but one of the many possible ways in which a new expression could be introduced and nothing more. The idea was not put forward even as the only possible way that a new *theoretical* term could be introduced. Other sorts of descriptions are certainly admissible. Thus Putnam writes: “I do not claim that a physical magnitude term can only be introduced by a causal description. But I do claim that a customary way of introducing a physical magnitude term is via a causal description...” (Putnam 1974, p. 176). It is an idea which is independent of the main point of the causal theory of reference, where the “causal” relation holds between different users of the term, and it is important to keep these two different ideas separated. And again, such an idea of causal descriptions is totally different from causal descriptivism.

Conclusions

It seems that most of the criticism of the new theory of reference in the philosophy of science is really directed towards a caricature of the theory. Not a single one of the key advocates of the new theory of reference have held such a view. Perhaps the causal theory of reference does not generalize as widely as some may have hoped. But in any case, it nevertheless opens interesting new possibilities in the theory of reference when compared to the more traditional descriptivist options.

One should, after all, remember what was our initial problem. Namely, according to Kuhn and Feyerabend, *all* expressions are theory-dependent to the extent that their meaning is exhausted by their theoretical context, and consequently, that there is no neutral observation language where the comparison of the rival theories is possible. Now it is sufficient for the comparison that *some* expressions function in another way. Therefore, if the causal account is adequate even for some terms, such as observational natural kind terms, this is sufficient to undermine the meaning-variance thesis, and the incommensurability thesis. The causal theory of reference, when correctly understood, can thus be an important ingredient in the realist toolkit for defending the rationality of science.⁴

⁴ This paper was presented in EPSA’07 in Madrid, which explains its compressed character. It is based on a much more extensive paper in progress.

References

- Bird, A. (1998). *Philosophy of Science*, London: Routledge.
- Bird, A. (2000). *Thomas Kuhn*, Chesman: Acumen.
- Boyd, R. (1973). "Realism, Underdetermination, and a Causal Theory of Evidence", *Noûs* 7, 1-12.
- Burgess, J. (2006). "Saul Kripke, Naming and Necessity", *Central Works in Philosophy*,
- Devitt, M. (1979). "Against incommensurability", *Australasian Journal of Philosophy* 57, 29–50.
- Devitt, M. (1981). *Designation*, New York: Columbia University Press.
- Devitt, M. and K. Sterelny (1987). *Language and Reality*, Basil Blackwell, Oxford.
- Devitt, M. and K. Sterelny (1999). *Language and Reality*, Second Edition, Blackwell, Oxford.
- Devitt, M. (1984/1999). *Realism and Truth*, First edition 1984, Second revised edition 1991, Oxford: Blackwell.
- Dupre, J. (1981). "Natural kinds and biological taxa", *Philosophical Review* 90, 66-90.
- Enc, B. (1976). "Reference of theoretical terms", *Noûs* 10, 261-282.
- Fine, A. (1975). "How to Compare Theories: Reference and Change", *Noûs* 9, 17-32.
- Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Science*. Cambridge: Cambridge University Press.
- Kitcher, P. (1978). "Theories, Theorists and Theoretical Change", *The Philosophical Review*, 87, 519-547.
- Kroon, F. (1985). "Theoretical terms and the causal view of reference", *Australasian Journal of Philosophy* 63, p. 143 – 166.
- Kroon, F. (1987). "Causal Descriptivism", *Australasian Journal of Philosophy* 65: 1–17.
- Lewis, D. (1984). "Putnam's Paradox", *Australasian Journal of Philosophy* 62: 221–36.
- Mellor, D. H. (1977). "Natural kinds", *British Journal of the Philosophy of Science* 28, 299-312.
- Newton-Smith, W. (1981). *The Rationality of Science*. London, Routledge.
- Niiniluoto, I. (1999). *Critical Scientific Realism*, Oxford University Press, Oxford & New York.
- Nola, R. (1980). "Fixing the reference of theoretical terms", *Philosophy of Science* 47, 505-531.
- Papineau, D. (1979). *Theory and Meaning*, Oxford: Clarendon Press.
- Psillos, S. (1999). *Scientific Realism. How Science Tracks Truth*. London: Routledge.
- Putnam, H. (1973). "Explanation and Reference", *Conceptual Change*, eds. G. Pearce and P. Maynard (Dordrecht: D. Reidel, 1973), pp. 199-221. Repr. in *Mind, Language and Reality* (1975), 196-214.
- Putnam, H. (1974). "The Refutation of Conventionalism", *Noûs* 8.1 (March 1974): 25-40. Revised version in *Semantics and Philosophy*, eds. M. Munitz and P. Unger (New York: New York University Press, 1974), 215-255. Repr. in *Mind, Language and Reality* (1975), 153-191.
- Putnam (1975a). "The Meaning of 'Meaning'." *Language, Mind and Knowledge. Minnesota Studies in the Philosophy of Science*, vol. 7, ed. K. Gunderson (Minneapolis: University of Minnesota Press, 1975), 131-193. Repr. in *Mind, Language and Reality* (1975), 215-271.
- Putnam, H. (1975b). "Language and Reality", in *Mind, Language and Reality* (1975), 272-290.
- Putnam, H. (1986). "Meaning Holism." *The Philosophy of W. V. Quine*, eds. L.E. Hahn and P.A. Schilpp (La Salle, Ill.: Open Court), 405-431.

- Putnam, H. (1988). *Representation and Reality*. Cambridge, Mass.: MIT Press.
- Raatikainen, P. (2006). "Against causal descriptivism", *Mind and Society* Vol. 5, No 1, 78-84.
- Sankey, H. (1994) *The Incommensurability Thesis*, Aldershot, Avebury.
- Sankey, H. (1997). "Incommensurability: The Current State of Play", *Theoria*, (12), 425–445.