Saving the intuitions: polylithic reference

Ioannis Votsis

Abstract   My main aim in this paper is to clarify the concepts of referential success and of referential continuity that are so crucial to the scientific realism debate. I start by considering the three dominant theories of reference and the intuitions that motivate each of them. Since several intuitions cited in support of one theory conflict with intuitions cited in support of another something has to give way. The traditional policy has been to reject all intuitions that clash with a chosen theory. A more radical policy, tied to some experimental philosophers, has called for the rejection of any evidential role for intuitions. I explore a largely ignored third alternative, i.e. saving intuitions (and their evidential role) even when they are at odds. To accommodate conflicting intuitions different sets of internally consistent (yet externally inconsistent) intuitions are taken to lend credence to different concepts of reference. In the current context, this means that the concepts of referential success and referential continuity are not monolithic. They are what I call 'polylithic'. This paper is as much about meta-philosophical concerns with the role of intuitions as it is about reference and the scientific realism debate. Regarding the former I hope that a blueprint will emerge for similar projects in other philosophical domains. Regarding the latter, I hope that polylithicity helps disentangle claims about referential success and continuity in the scientific realism debate by making perspicuous which concepts are best equipped to evaluate the realist's epistemic claims against the historical record of science.

Keywords   Scientific realism · Reference · Theory change · Intuitions · Experimental philosophy

I. Votsis (✉)
Department of Philosophy, University of Duesseldorf, Geb. 23.21/04.86, Universitaetsstrasse 1,
40225 Duesseldorf, Germany
e-mail: votsis@phil-fak.uni-duesseldorf.de

1 Introduction

Most scientific realists nowadays would endorse an argument like the following: The empirical and explanatory success of theories or theory-parts is a good indicator of their approximate truth. In turn, approximate truth is a good indicator of referential success.[1] Successor theories typically preserve all of the empirical and explanatory success of their predecessors as well as add to it. They are thus in general strictly more approximately true than their predecessors. Moreover, by preserving their predecessors' approximately true parts they preserve any referential success the predecessors enjoy. This implies that successor theories that are more approximately true than their predecessors are typically also referentially continuous with them.

An evaluation of these claims requires a clear grasp of the concepts involved. My aim in this paper is to clarify the concepts of referential success and of referential continuity. I start by considering the three dominant theories of reference, namely descriptivist, causal-historical and causal-descriptivist theories, and the intuitions that motivate each of them. Since several intuitions cited in support of one theory conflict with intuitions cited in support of another something has to give way. Two policies have thus far proved popular. The traditional policy has been to reject all intuitions that clash with a chosen theory. A more radical policy, tied to some experimental philosophers, has called for the rejection of any evidential role for intuitions. I explore a largely ignored third alternative, i.e. saving intuitions (and their evidential role) even when they are at odds. To accommodate conflicting intuitions I perform a compartmentalisation manoeuvre whereby different sets of internally consistent (yet externally inconsistent) intuitions lend credence to different concepts of reference. I argue that so long as we identify which concept is employed in which circumstances, some disputes about reference disappear. What is more, I illustrate how different concepts can be used to make sense of the historical record of science and to evaluate scientific realist claims.

A few stage-setting remarks are in order. First, most discussions of reference concern everyday language term reference and in particular the reference of proper names. Even though I will be drawing substantially on these discussions, as many other philosophers of science have done and continue to do, my primary focus will be on scientific term reference. By and large, material presented prior to Sect. 6 concerns both ordinary term as well as scientific term reference. Second, I do not intend to offer a comprehensive survey of theories of reference. I merely wish to concentrate on the most prominent versions as well as the most conspicuous supporting and opposing intuitions. Third, I do not intend to defend a fully developed view of the nature of intuitions. Instead, the various intuitions presented in this paper are expressed so as to conform to three desiderata: (i) they ought to reflect the internally uniform practices of different groups of competent language users, (ii) they ought to avoid philosophically loaded terms and (iii) they ought to be rationally evaluable. To comply with

---

[1] Inductive inferences from approximate truth to referential success turn out to be erroneous less often than they do the other way around. Indeed, if the correspondence theory of truth is correct it seems that the correctness of inferences from approximate truth to referential success is guaranteed since truth (and presumably even approximate truth) presupposes referential success but not vice-versa.

the first desideratum I attempt to narrow in on those intuitions members of a group who make similar judgments with respect to the application of a given concept have in common. To comply with the second desideratum I obviate philosophically loaded terms like 'reference' and 'object' in favour of more neutral terms like 'talking about' and 'thing'. To comply with the third desideratum I present the relevant intuitions in propositional form. Although in their natural state most intuitions are probably too obscure to possess a propositional form, we can reasonably reconstruct whatever is propositional about them or at least whatever propositions they are meant to prop up. The intuitions presented below are thus not necessarily actual intuitions but they are at least sensible proxies. Fourth, I only offer a rather small sample of the available intuitions. Fifth, it is worth keeping in mind that there is always some evidential distance between actual intuitions and philosophical theories. At best actual intuitions inductively support the intuitions as I reformulate them and they in turn inductively support the coveted philosophical theories.

2 Descriptivism

The core idea in descriptivist theories is that reference is fixed by virtue of a term's associated descriptions. Notable proponents include Frege (1892/1997), Russell (1905) and Searle (1958).[2] Let us start with a rather simplistic formulation that perhaps nobody ever advocated:

Def. 1: A term t refers to an entity a if and only if a satisfies all the descriptive claims associated with t.[3]

Since the first definition applies only to object terms, we need another one for predicates:

Def. 2: A term t refers to a property X if and only if any object with property X satisfies all the descriptive claims associated with t.[4,5]

For expediency let us hereafter forgo separate definitions for object and predicate terms, unifying the two as follows:

Def. 3: A term t refers to a(n) entity a (/property X) if and only if a(/any object with property X) satisfies all the descriptive claims associated with t.

Descriptivism is in tune with a number of widespread intuitions about reference. One such intuition the theory (as it is presented above) is meant to draw support from and hence satisfy is the following:

[2] In the early days of descriptivism reference was thought to be fixed by all and only those descriptions which analytically explicate the meaning of the referring term. Quine's well-known critique of the analytic-synthetic distinction precipitated the emergence of forms of descriptivism that dropped the analytic explication requirement. In what follows, I only consider examples of non-analytic descriptivist theories.

[3] Please note that definitions 1–6 do not discriminate between ordinary and scientific terms.

[4] Henceforth, properties are construed broadly to include relations.

[5] One need not assume that predicates refer directly to properties but can instead opt for the view that predicates refer to the set of all objects with a given property or even to each individual object in that set (Devitt 2003, p. 904).

(a)   To successfully talk about a thing requires that all of our ideas about it hold.

There is a certain naturalness to (a) since we sometimes tend to associate successful talk about an object with our ability to correctly attribute properties and relations to it that it does indeed possess. Suppose I am attempting to refer to Flavio, one of my niece's cats. Were I to mistakenly describe Flavio as white with brownish spots, my niece would protest that I am not talking about Flavio. The suitability of my niece's judgment can be maintained even when all my other Flavio-attributions are true. After all, she believes that one incorrect attribution is sufficient to foil the proper identification of her cat.

A related intuition that the above descriptivist theory is meant to satisfy can likewise be illustrated via the disagreement between interlocutors. The difference is that this time the focus is not on the incorrectness of one interlocutor's attribution but rather the mere disagreement itself:

(b)   If yours and my ideas about some thing are not in agreement then we are not talking about exactly the same thing.

Who hasn't been in a situation where a discrepancy in the descriptions used by two or more speakers elicits the remark 'They are surely not talking about the same thing!'[6]

Intuitions (a) and (b) are congenial to, though not exclusively associated with, the further intuition:

(c)   No idea a person may have fails to be about some thing.

Essentially this means that every description corresponds to some object. Though radical sounding at first, there is certainly a sense in which (c) reflects some folk usage. Competent language users often attribute thinghood merely on a stipulative basis. For example, people talk about counterfactual things like the child two celebrities would have were they to reproduce and even talk about impossible things like the round square. Thus at least some speakers are willing to be maximally charitable in their ascription of referents. Of course, that's not to say that competent language speakers have the same ontic and epistemic commitments to such things as they do to real things.

Let us turn now our attention to three related problems. The *first problem* is that descriptivist theories are too demanding. Lakatos once famously pointed out that scientific theories are born refuted (1978, p. 5). In the current discussion that translates as the assertion that at least some of the descriptions associated with scientific terms are false. For descriptivists who maintain that successful reference requires the satisfaction of all associated descriptions this means that no scientific term refers.[7] By *modus tollens* such a descriptivist view, branded by its critics as 'naïve descriptivism', is patently false if one believes, as most realists do, that at least some scientific terms refer.

---

[6] With minor adjustments intuitions (a) and (b) are in fact logically equivalent. In a longer version of this paper I included the proof for this equivalence.

[7] There are even stronger ways of reading intuition (a), i.e. taking the clause 'requires that all our ideas about that thing hold' to mean that *all and only* our ideas about that thing hold.

One obvious reply that has been proposed is to lower the requirements of successful reference. Here's a generic account of a modified descriptivist theory:

Def. 4: A term *t* refers to a(n) entity *a* (/property *X*) if and only if *a* (/any object with property *X*) satisfies a certain special subset of the descriptions associated with *t*.

This definition immediately gives rise to another problem. I call this *second problem* the 'ntity-lity problem', for it concerns the qua-ntity and qua-lity of the descriptions involved. How many and/or which descriptions are enough? Otherwise put, how do we define the special subset of descriptions required to establish reference?

That all of our ideas about a thing hold does not always mean that they are sufficient to uniquely identify it. We may call this a case of 'underdetermination'. Two mutually exclusive options are available. Either we deem unique identification necessary for successful reference or else we deem it unnecessary. The converse condition, i.e. overdetermination, not only occurs but quite a few philosophers would argue it is the norm. It happens when a proper subset of our true ideas about a thing appears to be sufficient to uniquely identify it. We also face a dilemma here. Either we deem that unique identification is sufficient for reference or else we deem it insufficient.

Some proposals have been (or can be) put forth to tackle the second problem. For example, Searle (ibid.) proposes a cluster view of proper names, according to which a name refers when "a sufficient but so far unspecified number of [descriptions] are true of [the given] object" (p. 171). Another proposal seeks to identify the special subset with those descriptions of the objects or properties that arise in the context of a mature science or theory. A related proposal appeals to essential or natural kind properties. The thought here is that some descriptions are more critical than others in determining the essence or nature of a particular object or property.

This brings us to the *third problem* which questions whether we have independent reasons to accept each of the proposed solutions to the second problem. That is, it asks if the conditions suggested to identify the special subset of descriptions are ad hoc. Take, for example, the cluster view proposal. If one of our aims is to find a theory of reference that does not require all the associated descriptions to be satisfied, merely codifying this aim in a view that denies naïve descriptivism is not satisfactory. Independent reasons are required for the adoption of the cluster view and such reasons would have to answer, among other things, why it is that some sets of descriptions are sufficient to fix reference.

In spite of these obstacles, there is no a-priori reason why a sophisticated descriptivist account could not possibly do justice to the ntity-lity problem without falling prey to the problem of ad-hocness. Any such account would presumably be motivated by some suitably modified version of intuitions (a)–(c). Having said this, there are intuitions that, at least prima facie, seem incongruent with all descriptivist theories. It is to such intuitions we next turn towards and in particular to those that support the causal-historical theory of reference.

3 Causal-historicism

The underlying idea in causal-historicism is that reference is fixed via a term's causal history. The view was famously proposed by Kripke (1972/1981) and developed further

by Putnam (1975). It has since acquired a great number of supporters including Boyd (1993) and Devitt (1990). Consider the following generic formulation:

> Def. 5: A term t refers to a(n) entity a (/property X) if and only if t was used to originally dub a(/X) after some causal contact with a(/X) or samples of a (/instantiations of X) and any speaker's subsequent use of t is causally linked back to the original dubbing.

The causal component plays a dual role. First, it requires that the speaker who initially baptised the entity or property have causal contact with it. For example, the baptiser must have interacted with or at least observed the given entity or property. Second, any speaker who subsequently employs that term to successfully refer must, however indirectly, be causally connected to the baptiser. In other words, such speakers learnt to use the term from causal contact with others who used it in the same way in a historical line that stretches back to the baptiser.

Consider the kinds of intuitions the causal-historical view is intended to draw support from.

(d)   Successful talk about a thing can occur even when all of our ideas about it do not hold.

One camp's intuitions in favour are the other camp's intuitions against. Intuition (d) contradicts intuition (a). Even before the first explicit formulation of a causal-historical theory, philosophers of language were keen to point out cases where successful reference does not require correct descriptive content. Donnellan (1966, pp. 364–365) uses the example of a person at a party who inquires about 'the man drinking a martini'. The definite description is successfully used referentially, according to Donnellan, even though the person is drinking water rather than a martini.

(e)   Two or more people may talk about the same thing even if they have conflicting ideas about it.

This intuition contradicts intuition (b). Think of Donnellan's example again but suppose now that even though the discussants all have conflicting ideas about the man drinking water at the party, they successfully identify, and refer to, him via ostension. Notice that for this intuition to hold it need not be the case that there is at least one individual whose ideas are correct, i.e. they could all be mistaken.

At this point one might expect the formulation of an intuition that contradicts (c). Yet causal theorists can take (c) on board. So long as causal contact determines reference it does not matter whether descriptions always correspond to some object, real or fictional. That being said, what we need is an intuition to underwrite the causal component of the theory.

(f)   To talk about some thing typically requires some sort of direct contact or at least indirect contact—i.e. contact with a person who has direct contact or who belongs in a chain of persons whose last link at least has direct contact—with that thing.[8]

---

[8] This intuition is logically independent from intuitions (d) and (e).

Consider an example. The term 'berkelium' denotes a radioactive metallic element with atomic number 97. The person who presumably dubbed it, its discoverer and Nobel laureate Glenn T. Seaborg, arguably had some causal contact with berkelium since he headed the cyclotron experiments that synthesised the element by bombarding americium with alpha particles. It seems right then to say that when Seaborg employed the term 'berkelium' he successfully referred to the element berkelium. A nuclear chemist who does not have access to a sample of berkelium can still correctly refer to the substance or instances of the substance by borrowing the term from others who do have some causal contact with a sample.[9] Finally, it seems correct to hold that someone who merely happens to utter the term 'berkelium' without prior direct or indirect contact cannot be successfully referring to the substance or any of its instances.

Like descriptivists, causal theorists come up against several problems. We can again identify three related and widely discussed ones. The *first problem* is that the causal theory is too liberal. It allows any old term to refer to some entity or property. The typical example given is that of the notion of dephlogisticated air. Nowadays we do not believe in the existence of phlogiston but since scientists in the eighteenth century used the term 'dephlogisticated air' in virtually all the cases where causal contact with oxygen was made, the causal theorist is forced to claim that 'dephlogisticated air' was in fact referring to oxygen all along.[10] For many philosophers of science these reconstructions of history are a bitter pill that they do not consider sensible to swallow.

To pre-empt such objections Putnam (1978) introduced the principle of the benefit of doubt. If the descriptions associated with an old theoretical term do not diverge unreasonably from those of its modern day counterpart, the principle allows us to brand the old theoretical term referential. That descriptions were not entirely absent from causal-historical theories is something that has been missed by several commentators. Kripke (1972/1981, p. 79) made it clear that the dubbing event may include descriptions and that it may even be solely based on them. In support of the latter scenario he offered the celebrated discovery of Neptune. Initially scientists had no ostensive (causal) contact with Neptune but fixed its reference merely on the basis of theoretical calculations, i.e. on the basis of descriptions.

If we allow descriptions to creep in then we again come face to face with the ntity-lity problem. This marks the *second problem* for the causal-historical theory. It is worth noting that the principle of the benefit of doubt is tactically identical to the modified versions of descriptivism discussed earlier. It aims to find the right balance between requiring that all descriptions must be satisfied and that none of them need to. In this respect, one may understand causal-historical theories as augmented descriptivist theories. The crucial question once more is when are we justifiably charitable? Unless some guidance is given opinions are likely to vary. Indeed, some philosophers have complained that the principle of charity trivialises the notions of referential success and continuity (e.g. Worrall 1994).

The causal theorists' answers to the ntity-lity problem mirror those given by the descriptivists, e.g. appeal is made to essential or natural kind properties, etc. For this

---

[9] Kripke adopts the idea of borrowing a reference from Strawson (1959, p. 182).

[10] For a congenial account see Schurz's contribution to this issue of Synthese.

reason, ad-hocness worries arise here also. This marks the *third problem* for the causal-historical theory. Independent reasons are required to warrant the application of the principle of the benefit of doubt, reasons that make it clear why in a given context it diagnoses referential success or failure, referential continuity or discontinuity. As with descriptivist theories, it is worth reminding that the obstacles cited above do not constitute an a-priori reason against a sophisticated causal-historical account.

4 Causal-descriptivism

As its name suggests, causal-descriptivism aims to fix reference by appealing to a combined strategy. The view can be traced back to Evans (1973). Other notable proponents include Lewis (1984), Kroon (1987) and Nola (1980). Evans takes reference to be fixed by the dominant (causal) source of the speaker's descriptions. Take the term 'aeroplane'. Since the main causal source of the descriptions we associate with this term, e.g. 'has a jet or a propeller engine', 'flies', etc., is aeroplanes, the term refers to aeroplanes. Expressed formally a causal-descriptivist theory might look something like this:

Def. 6: A term *t* refers to a(n) entity *a* (/property *X*) if and only if the dominant (causal) source of any descriptive content associated with *t* is *a*(/*X*).

Following descriptivist theories, the causal-descriptivist view complies with the general intuition that descriptions are necessary and sufficient for successful reference. Following causal-historical theories, the view satisfies the general intuition that causation is a necessary component of reference fixing. That descriptions are necessary and sufficient for successful reference can be reconciled with the idea that causation is a necessary component by pointing out that the latter acts as a filter on the former. If you like, we determine the special subset of descriptions by appeal to those descriptions that encode what is causally relevant.

An instructive example from Evans about ordinary term reference will help make clear how a causal-descriptivist theory is supposed to function. Suppose we discover an urn which contains various mathematical proofs. Since the urn is inscribed with the name 'Ibn Khan', it is widely assumed thereafter that credit for the proofs goes to the person with that name. Suppose further that Ibn Kahn was merely the person who transcribed the proofs many years after their creation. What does the term 'Ibn Kahn' refer to? The scribe? The mathematician who constructed the proofs? A fictional person? No person whatsoever? The descriptive theory of reference might rule in favour of any of these answers depending on what its advocates take to be the special subset of descriptions that needs to be satisfied. In the case of the causal-theory of reference the situation is also fuzzy. For example, if the initial baptism and subsequent usage have already established a consensus amongst the ancients that 'Ibn Kahn' denotes the scribe then the causal-theorist will be obligated to follow suit. If, on the other hand, a more refined causal theorist insists that a new baptism takes place upon the urn's modern-day discovery, then the term 'Ibn Khan' as it is used by members of the mathematical community today denotes the author of those proofs. In Evans' view, modern-day mathematicians never intended to refer to the scribe but rather to

the author of the proofs. After all, the author is the 'dominant causal origin' of the most relevant description associated with the name, i.e. 'the person who constructed the proofs'.

The causal component plays a different role here than it does in causal-historical theories. Whereas in those theories the causal component relates the baptiser (and the baptism term) to the referent as well as to subsequent users, in causal-descriptivist theories the causal component relates the descriptive content to the referent. To complicate matters even further, in some theories the causal component plays yet another role. For example, in Stathis Psillos' causal-descriptivist theory that we shall shortly be examining the term 'causal' qualifies the properties the relevant descriptions are about. In other words, Psillos' theory focuses on descriptions of causal properties as opposed to Evans' which focuses on the main causal source of a set of descriptions.

Let us put together a couple of potential intuitions some causal-descriptivist theories are intended to satisfy:

(g)   To successfully talk about a thing requires that our ideas primarily originate from contact with that thing.

(h)   Two or more people may talk about the same thing so long as any non-conflicting ideas they have primarily originate from contact with that thing.

Just like intuitions (a)–(f), (g) and (h) can be sharpened so as to reflect a whole range of intricacies that their advocates would like to incorporate.

For obvious reasons, the ntity-lity and ad-hocness problems make their appearance here also. Instead of hammering the same nail again, I will consider some difficulties with Psillos' version of the causal-descriptivist theory, which is tailor-made for the scientific realism debate. He explains his view as follows: "A term t refers to an entity x if and only if x satisfies the core causal description associated with t" (1999, p. 296). The core causal description associated with a term is the description of "the [kind-constitutive] properties by virtue of which it [i.e. a posited entity] plays its causal role vis-à-vis the [relevant] set of phenomena" (p. 295). What are kind-constitutive properties? They "are those whose presence in an item makes that item belong to a kind" (p. 288). In order to make explicit the role of his theory in adjudicating referential disputes in the scientific realism debate, Psillos introduces a notion of *referential continuity* which he takes to piggy-bag on his notion of successful reference:

> Two terms t' and t denote the same entity if and only if [i] their putative referents play the same causal role with respect to a network of phenomena; and [ii] the core causal description of t' takes up the kind-constitutive properties of the core causal description association with t. (Psillos 1999, p. 296)

In other words, for two successive terms to refer to the same entity, the historically later term must inherit the core causal description associated with the historically former term and the two must concern the same domain of phenomena. Psillos cites the term 'ether' as one of several examples that fit his approach. He takes users of the term 'ether' in the nineteenth century to be successfully referring to the electromagnetic field.

Several problems afflict Psillos' theory. The first problem concerns Psillos' two definitions, which, as it turns out, are inconsistent. Consider the following example. Suppose that James Ladyman, the philosopher of science, has two kind-constitutive properties, creature with a gentle spirit and creature with a fist of steel. Suppose further that we have two successive theories about James with terms t and t' respectively. The core causal description associated with term t' contains only 'creature with a gentle spirit' and the core causal description associated with term t contains only 'creature with a fist of steel'. Now according to the definition of successful reference both t' and t refer to James, as it seems they should. However, according to the definition of referential continuity t' and t cannot both be referring to James since t' does not take up the kind-constitutive properties of the core causal description associated with t. The problem stems from the fact that in his attempt to formulate a notion of referential continuity Psillos inadvertently redefines the notion of referential success by stating conditions under which terms denote an entity.

Psillos brings up a second serious objection to his theory when he says "But, an objector may ask, how (and when) is the core description to be singled out?" (p. 297). I take this to be tantamount to the ntity-lity problem. Psillos attempts to dismiss the problem by reiterating what many other causal and causal-descriptivist theorists have said before him, namely that "some descriptions associated with a term are less fundamental in view of the fact that the posited entity would play its intended causal role even if they were not true" (ibid.). Though a sensible thing to say, it still doesn't give us a non-ad hoc and unambiguous way to determine the fundamental from the non-fundamental descriptions. We might, for instance, contest Psillos' account of the ether's core causal description, by arguing that several of its fundamental properties, e.g. that ether molecules oscillate and that transverse waves require a solid medium, were not taken up by the mature electromagnetic field's core causal description (Stanford 2003).

As with the other theories of reference, I am not arguing here that Psillos' theory or causal-descriptivist theories in general are a lost cause, unamenable to successful modification. I am simply pointing out some of their limitations as well as the intuitions they are meant to satisfy.

5 Saving the intuitions

The realisation that various intuitions are conflicting leads to the following trilemma: Either we (1) identify one set of internally consistent intuitions that we take to count as evidence for one theory of reference or (2) reject any evidential role for intuitions or (3) maintain that a number of internally consistent (but externally inconsistent) sets of intuitions count as evidence for their respective theories and corresponding concepts. The first option is taken up by the majority of supporters of the above three theories. Each group of supporters deem their own intuitions to be superior and on that basis discard the other theories and their associated intuitions.[11] The second option is a

---

[11] It may be that in practice theories are chosen first and intuitions are chosen as a consequence. Where this happens it is difficult and perhaps impossible to maintain that the relevant intuitions are evidence for that theory.

relatively recent development in meta-philosophical discussions about methodology. Some naturalists and experimental philosophers (e.g. Cummins 1999) have rejected the practice of employing intuitions to justify philosophical theories. According to them, only empirical evidence is equipped for that job. I aim to explore the third option, namely that each set of internally consistent intuitions lends credence to a different concept. In the current context, this means that the concepts of referential success and referential continuity are not monolithic. They are what I call 'polylithic'.

Are there really so many concepts in use? To answer this question we must first ask if there are many intuitions in use. On the basis of the available empirical evidence (e.g. Machery et al. 2004), there seems to be a prima facie strong case for the relative multiplicity of intuitions. These may vary from culture to culture, from individual to individual and even from time to time in the same individual. If intuitions genuinely guide at least some of our referential judgements then they strongly suggest that our linguistic practices are peppered with a plethora of referential concepts.[12] The upshot of adopting the third option in the trilemma is therefore a more accurate portrayal of these practices. One and the same term is often used at least somewhat differently by different individuals and may even be used differently by the same individual over time. Unless we opt to eliminate *all* conflicting uses as incorrect applications of one and the same concept, it is reasonable to assume that there are numerous concepts at play. Such concepts are not unrelated since their extension tends to overlap significantly. Think of the three theories of reference and the corresponding concepts they propose. Their rulings on referential success diverge only on the fringe cases. This is not surprising since they all aim to incorporate as large a catalogue of common-sense cases of referential success and failure as possible. Indeed we can explain why communication does not constantly break down precisely because of the substantial extensional overlap these concepts tend to enjoy. Conversely we can explain the relatively few occasions when it does break down by pointing out those elements of their extensions that do not coincide.

Ought there be a plurality of referential concepts? Yes, but that does not invite a free-for-all! Linguistic practice reveals several different concepts, standards of concept application and intuitions but not all of them are worth considering. We at least want to weed out those that fail to meet minimal logical and/or rationality criteria. Conversely, we need to plumb for concepts, standards of concept application and intuitions that have hitherto remained undetected. Some of them will be the product of scientific cultivation. For example, a modern-day physicist's intuitions about space and time will be very different from those of a physicist living in the nineteenth century.[13] In short, we want our concepts to exhibit the 'best' current practices but also the practices that best extend the current ones. Acknowledging a plurality of concepts does not mean that certain concepts are not more natural or better suited than others for partic-

---

[12] Strictly speaking it does not matter to my argument whether there are many referential concepts or many different instantiations of the same concept so long as each instantiation is regulated by its own standards of correctness.

[13] Some concepts, especially those in science and mathematics, are so technical, well-defined and accurate that after some point in time we may no longer need to consider variants of them. In such cases, we may still speak of monolithic concepts.

ular tasks. In the context of the scientific realism debate certain notions of referential success and continuity will turn out to be much more useful than others. For instance, referential continuity concepts will need to be sufficiently stringent so as to not trivialise continuity and sufficiently supple so as to not preclude the possibility of continuity. So long as we are aware which concepts are used under what circumstances, we will be able to solve and even dissolve various disputes.

A word of caution before we proceed to the next section. Each option in the trilemma is not as insular as I have hitherto portrayed. The first option is not only available to traditional supporters of the three theories of reference who primarily employ conceptual analysis to practice their trade. It is also available to moderate experimental philosophers and naturalists who resist the temptation to reject the evidential role of intuitions. Instead such philosophers argue that we should vet intuitions on the basis of experimental studies (see, for example, Nahmias et al. 2005).[14] The second option can also be unraveled in different ways. For example, those who deny that intuitions play any evidential role can choose between a monolithic and a polylithic understanding of concepts. The only difference between a second-option polylithist and a third-option polylithist is the way each motivates their polylithism—in the one case without an appeal to intuitions and in the other with such an appeal. Finally, the third option is compatible with both moderate experimental and non-experimental approaches to philosophy. What is more, one may choose between saving all concepts and intuitions and weeding out some as undesirable—I have suggested that the latter option is more prudent. What matters most to those who advocate the third option is that inconsistent intuitions are compartmentalised into separate internally consistent sets each of which is evidentially relevant to a different concept.

6 Polylithic reference[15]

In this section I illustrate how different concepts of referential success and continuity can be used to make sense of the historical record of science and to evaluate scientific realist claims. I will not present an exhaustive list of reference concepts that competent language users possess or would benefit from possessing. There are simply too many such concepts so I will restrict my attention to a selected few. For the same reasons, I will not attempt to catalogue all the reference concepts that are significant for scientific realism. I will, however, explore at least one concept whose satisfaction demands are in sync with the epistemic claims current scientific realists make. Finally, I will not test the concepts below against possible counterexamples or against long-standing problems like ntity-lity and ad-hocness. Rather I will try to indicate the usefulness of

---

[14] Alexander and Weinberg (2007, pp. 62–63), call the more radical experimental philosophers who want to restrict (partly or fully) the evidential use of intuitions 'restrictionists', and those who want to select intuitions on the basis of experimental results 'proper foundationists'.

[15] Polylithicity is sometimes considered as a limited-range fall-back position by non-experimental philosophers who are forced to accept the non-universality of intuitions in certain domains. To the best of my knowledge no account of reference proposed up to now is genuinely polylithic. Kitcher's (1993) token-type account of reference comes closest.

having different concepts of reference in sorting out some disputes in the scientific realism debate.

We start with a notion of reference that is particularly undemanding.

Def. 7: A scientific term t refers(MN) to a real entity a (/property X) if and only if t is used to consistently identify $a(/X)$ or at least to consistently identify the causal source of phenomena associated with $a(/X)$, either via a potentially truthful description or via some actual contact with the source.[16]

The motivation for this notion of reference is that it saves some versions of the aforesaid intuitions, e.g. not neglecting causal contact and successfully talking about an object that we have merely attempted to describe. Let us also define a corollary notion of referential continuity refcont(MN) derivatively:

Def. 8: Two terms t and t' are refcont(MN) if and only if they refer(MN) to the same entity or property.

The two definitions are strikingly easy to satisfy. To illustrate this, consider the following example. Aristotle's expression 'tending towards its natural place' refers(MN) to gravity for its corresponding notion was postulated to explain, among other things, phenomena relating to falling objects.[17] The mere attempt to describe such phenomena is sufficient for reference(MN) but so is the causal contact Aristotle and his contemporaries had with falling objects. Apples fell then just as they did during Newton's time and continue to do so today. Since such phenomena are associated with our modern understanding of gravity, the expression 'tending towards its natural place' is refcont(MN) with the expression 'gravity as spacetime curvature' (for a congenial account, albeit one that's used for different purposes, see Friedman 2001). Most, if not all, scientific realists would consider this kind of referential continuity extremely feeble since Aristotle and his supporters had nothing in mind that is remotely similar to the general theory of relativity. Nonetheless, Aristotle, Einstein and supporters of their theories certainly share an interest in explaining phenomena relating to gravity and in this minimal respect the two expressions and the corresponding theories are continuous.

That the two definitions are so easy to satisfy does not mean that they are trivially satisfiable. To see this consider a case of referential failure. The term 'celestial sphere' does not refer(MN) to gluon particles. The celestial spheres were postulated to explain celestial phenomena like the daily motion of the 'fixed' stars but not the kind of phenomena we associate with gluons, namely that they mediate strong (colour) interactions between quarks, bind protons and neutrons in atomic nuclei, etc. A fortiori, we can say that the term 'celestial sphere' is not refcont(MN) with the term 'gluon'. Note

---

[16] Even though most discussed cases in the history of science involve the frequent use of t by a group of individuals to consistently identify an object or property, the definition is intentionally slack to allow cases where the term is infrequently used in that capacity. Ditto for all the definitions given below.

[17] According to Aristotelian physics sublunar bodies like rocks tend to move towards the centre of the earth. Yet, the Aristotelian notion is also meant to explain phenomena related to rising objects, e.g. sublunar bodies like fire tend towards the heavens. This does not frustrate the Aristotelian expression's successful reference(MN) to gravity because def. 7 is sufficiently weak so as to allow expressions to refer to disjoint sets of phenomena.

also that there are contexts where refer(MN) is more difficult to satisfy than the naïve descriptivist theory that underwrites intuition (c) since the former prohibits reference to non-real objects.

Let us step up the satisfaction demands.

> Def. 9: A scientific term t refers(TD) to a real entity a (/property X) if and only if (i) t is used to consistently identify $a(/X)$ or at least to consistently identify the causal source of phenomena associated with $a(/X)$ and (ii) some of the (non-trivial) theoretical descriptive claims associated with t are true of $a(/X)$.

This notion of reference saves intuitions about not neglecting causal contact and about successfully talking about an object despite having some false associated ideas. It also saves more philosophically refined intuitions like the demand for the satisfaction of one or more theoretical descriptions. As before we define a corollary notion of referential continuity derivatively:

> Def. 10: Two successive scientific terms t and t' are refcont(TD) if and only if they refer(TD) to the same entity or property.

Henceforth t' is the successor term and t the predecessor.

Obviously this notion is harder to satisfy. Terms that satisfy it enjoy a thicker kind of referential continuity that is capable of tackling some of the historical objections faced by the realist. Consider an example of referential success and referential continuity. The term 'caloric' in the early nineteenth century refers(TD) to, and is refcont(TD) with, heat as it is understood in classical and statistical thermodynamics. This is the case because certain of the theoretical descriptions associated with the term 'caloric', e.g. Sadi Carnot's principle of maximum efficiency, are still thought to be true of heat systems today.[18] The gravity example given earlier now serves as an example of referential failure and referential discontinuity. The expression 'tending towards its natural place' neither refers(TD) to gravity nor is refcont(TD) with the expression 'gravity as spacetime curvature' since it not associated with any (non-trivial) theoretical descriptive claims that are true of gravity.

There is a stronger sense of continuity that scientists and scientific realists aspire towards. It occurs when a successor theory's term inherits all of the non-trivially true theoretical and empirical descriptions that the predecessor theory associated with a corresponding term. This is roughly the kind of continuity that Psillos' causal-descriptivism attempted to secure. Let us encapsulate the desired continuity semi-derivatively as follows:

> Def. 11: Two successive scientific terms t and t' are refcont(TD)* if and only if (i) they refer(TD) to the same entity a (/property X) and (ii) t' inherits all of the (non-trivial) theoretical and empirical descriptive claims true of $a(/X)$ that are associated with t.

---

[18] Carnot's principle states that a heat engine operating between two reservoirs at different temperatures will have a maximum efficiency, i.e. given a certain input of heat there is a limit on how much of that heat can be converted into work. Indeed, it states that no actual engine can be perfectly efficient. That is achievable only by an *ideal* Carnot engine.

In effect, refcont(TD)* requires that those parts of a predecessor theory that are true of some object or property be preserved in the successor theory.[19] This bodes well with the lofty requirements current scientific realists place on a theory of reference. If, as was argued in the first paragraph of this paper, a successor theory is to be strictly more approximately true than a predecessor then all of the latter's true theoretical and empirical descriptive claims (including those relating to a particular term t) must be preserved in the successor theory. Take the earlier 'caloric' example. The term 'caloric' is arguably refcont(TD)* with heat as it is understood in classical and statistical thermodynamics since Carnot's principle of maximum efficiency, among other true theoretical and empirical descriptive claims, is preserved in those accounts in the form of the second law of thermodynamics.

Finally let us look at a notion whose satisfaction demands are perhaps impossibly high.

> Def. 12: A scientific term t refers(MX) to a real entity a (/property X) if and only if (i) t is used to consistently identify $a(/X)$ and (ii) all and only the theoretical and empirical descriptive claims associated with t are true of $a(/X)$.

Most probably no scientific term qualifies as referentially successful under these conditions. Nonetheless the notion saves exaggeratedly optimistic intuitions we might have about the indispensability of causal contact and about the virtues of acquiring all and only true descriptions of an object. Since referential continuity is understood as involving progress and terms satisfying refer(MX) cannot progress further, i.e. they cannot be associated with any more truths about their target objects, a suitable definition of refcont(MX) must require only that t' refers(MX). The remaining conditions for such a definition can vary according to the desired strength. Here's a sufficiently strong formulation of refcont(MX):

> Def. 13: Two successive scientific terms t and t' are refcont(MX)* if and only if (i) t' refers(MX) to $a(/X)$, (ii) t nearly refers(MX) $a(/X)$ and (iii) t' inherits all of the (non-trivial) theoretical and empirical descriptive claims true of $a(/X)$ that are associated with t.

The strength of the notion refcont(MX)* obviously depends in part on the strength of the notion 'nearly refers(MX)'.

The concepts discussed above are only showroom examples of the kinds of concepts that are required to satisfy the remarkable variety of intuitions. Refer(MN) and refer(MX) are intended to represent two extremes—not the only extreme ones—of the spectrum of noteworthy referential concepts. Naturally, many concepts with varying degrees of satisfaction demands fall in between. Most versions of the three dominant theories of reference as well as refer(TD) are amongst them. Refer(TD) is intended to represent the kind of concept that current scientific realists need in order to ascertain their epistemic claims. These claims can be tested against how well the historical record of science fits sufficiently demanding notions of reference and referential continuity. The stronger the epistemic claims a scientific realist makes the higher the

---

[19] Refcont(TD)* allows t' to inherit also false claims about a. Stronger formulations can be given which forbid this from happening.

satisfaction demands that need to be met. Otherwise put, the thicker the referential continuity established, the stronger the case for scientific realism.

The current paper sticks its neck out by making the following testable prediction: Other things being equal, the more empirically successful a particular theory becomes the more likely it is that its successor theories will satisfy increasingly stronger notions of referential success and continuity. Future scientific realists will probably need referential concepts stronger than refer(TD) and refcont(TD)* to ascertain their own particular epistemic claims since their challenges will presumably be more sophisticated—that is they will have higher satisfaction demands.

At this point readers might be unclear why we need a multitude of referential concepts if only one is required to test scientific realism against its opponents. There are at least two reasons for this. The first reason concerns the diachronic character of the debate. Since scientific realists and their opponents adapt their epistemic claims over time the requisite referential concepts adapt with them. That already implies a multiplicity of referential concepts each tied to a different set of satisfaction demands. The second reason concerns the synchronic character of the debate. Even from the perspective of a set of fixed challenges to scientific realism, theories and their posits can be referentially successful and continuous in different ways. So long as we are clear about what kind of success and continuity is at stake, we can avoid a lot of unnecessary disputes about reference. Labelling each kind with a unique concept allows us to do just that. As we saw earlier, we need not deny that Aristotle's expression 'tending towards its natural place' is referentially continuous to Einstein's 'gravity as spacetime curvature' in some sense, e.g. refcont(MN), to accept that it is referentially discontinuous in others, e.g. refcont(TD).

I would like to end this section by bringing to light one more complication. Following convention, I have up to now spoken about referential continuity as a one-one relation. That is not strictly speaking always the case. Sometimes two or more predecessor notions are replaced by a single successor notion. For example, both Kepler's notion of planetary motion and Galileo's notion of a freely falling body get replaced by Newton's notion of force. Sometimes the converse happens. A single predecessor notion is replaced by two or more successor ones. For example, the classical notion of kinetic energy is replaced by the quantum mechanical notion for sub-atomic objects and the relativistic notion for all other objects. This complexity needs to be reflected in our formulations of referential continuity concepts if they are to be fair to the historical record of science.

7 Conclusion

This paper was as much about meta-philosophical concerns with the role of intuitions as it was about theories of reference and the scientific realism debate. Regarding the former I hope that a blueprint has emerged for similar projects in other philosophical domains. Regarding the latter, I hope to have provided compelling reasons why an account of multiple referential concepts does justice to the motley of linguistic practices by providing a framework within which each practice and the intuitions that underlie it tallies with a different concept. Ultimately, I hope that such an account helps

disentangle claims about referential success and continuity in the scientific realism debate by making perspicuous which concepts are best equipped to evaluate the realist's epistemic claims against the historical record of science.

References

Alexander, J., & Weinberg, J. M. (2007). Analytic epistemology and experimental philosophy. *Philosophy Compass, 2*(1), 56–80.

Boyd, R. (1993). Metaphor and theory change: What is a 'metaphor' a metaphor for?. In A. Ortony (Ed.), *Metaphor and thought* (pp. 481–532). Cambridge: Cambridge University Press.

Cummins, R. (1999). Reflections on reflective equilibrium. In W. Ramsey & M. DePaul (Eds.), *Rethinking intuition* (pp. 113–127). New York: Rowman & Littlefield.

Devitt, M. (1990). Meanings just ain't in the head. In G. Boolos (Ed.), *Meaning and method: Essays in honor of Hilary Putnam* (pp. 79–104). Cambridge: Cambridge University Press.

Devitt, M. (2003). Theories of reference. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (Vol. 3, pp. 904–907). London: Nature Publishing Group.

Donnellan, K. (1966). Reference and definite descriptions. *Philosophical Review, 75*, 281–304.

Evans, G. (1973). The causal theory of names. *Proceedings of the Aristotelian Society, 47*, 187–208.

Frege, G. (1892/1997). On sinn and bedeutung. (Reprinted in M. Beaney (Ed.), *The Frege reader* (pp. 151–171). Oxford: Blackwell.)

Friedman, M. (2001). Transcendental philosophy and a priori knowledge: A neo-Kantian perspective. In P. Boghossian & C. Peacocke (Eds.), *New essays on the a priori* (pp. 367–384). Oxford: Oxford University Press.

Kitcher, P. (1993). *The advancement of science: Science without legend, objectivity without illusions*. Oxford: Oxford University Press.

Kripke, S. (1972/1981). *Naming and necessity*. Oxford: Blackwell

Kroon, F. (1987). Causal descriptivism. *Australasian Journal of Philosophy, 65*, 1–17.

Lakatos, I. (1978). Introduction: Science and pseudoscience. In G. Currie & J. Worrall (Eds.), *The methodology of scientific research programmes* (pp. 1–7). Cambridge: Cambridge University Press.

Lewis, D. (1984). Putnam's paradox. *Australasian Journal of Philosophy, 62*, 221–236.

Machery, E., et al. (2004). Semantics, cross-cultural style. *Cognition, 92*(3), B1–B12.

Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2005). Surveying free will: Folk intuitions about free will and moral responsibility. *Philosophical Psychology, 18*(5), 561–584.

Nola, R. (1980). Fixing the reference of theoretical terms. *Philosophy of Science, 45*, 505–531.

Psillos, S. (1999). *Scientific realism: How science tracks truth*. London: Routledge.

Putnam, H. (1975). *Mind, language, and reality*. Cambridge: Cambridge University Press.

Putnam, H. (1978). *Meaning and the moral sciences*. London: Routledge and Kegan Paul.

Russell, B. (1905). On denoting. *Mind, 14*, 479–493.

Schurz, G. (forthcoming). When empirical success implies theoretical reference. *Synthese*.

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

Stanford, K. (2003). Pyrrhic victories for scientific realism. *Journal of Philosophy, 100*, 553–572.

Strawson, P. (1959). *Individuals: An essay in descriptive metaphysics*. London: Methuen.

Worral, J. (1994). How to remain (reasonably) optimistic: Scientific realism and the 'luminiferous ether'. In D. Hull, M. Forbes & R. M. Burian (Eds.), *PSA 1994* (pp. 334–342). East Lansing, MI: Philosophy of Science Association.