CHAPTER 10

Causal descriptivism and the reference of theoretical terms

Stathis Psillos

1 INTRODUCTION

An adequate theory of reference of theoretical terms should satisfy two important conditions.

I. The burden of reference of theoretical terms rests with theory in the sense that what they refer to is determined (at least to a large extent) by the theory in which they feature prominently.
II. Theoretical terms are transtheoretical in the sense that they can refer to the same entity even though they may occur in different theories (or, similarly, different terms may refer to the same entity).

There are many reasons why this should be so. When it comes to (I), it should be obvious that unlike many ordinary objects, theoretical entities cannot be pointed at, perceived by the naked eye, presented to our sensory modalities, and the like. If anything, we have causal contact with them by virtue of their effects; and cognitive contact with them, by virtue either of our causal contact with them or of the confirmation of theories that are about them (or both). Even though causal contact with theoretical entities is not theoretical *in itself*, ascertaining it requires or relies upon theoretical knowledge (or at least beliefs), since even if the causal relation is not itself theoretical, what causes certain effects (and hence what it is that we have causal and cognitive access to) should be theoretically identified by its properties. When it comes to (II), unless we take seriously the possibility that each and every theory that emerges presents us with an image of the world, fresh and totally unrelated to whatever has gone on before, there must be referential continuity between succeeding theories as a necessary condition for being able to talk about the same entities, even though different things may be said of them by different theories. Condition (II) is necessary for progress, for blocking incommensurability, and for developing a fairly realist image of the world.

It might be protested, right from the outset, that these conditions are tailored to a realist approach to science. This is only partly true. They are indeed significant for anyone who takes scientific theories literally – for semantic realists, let us say. Clearly, those who take it that theoretical terms are cognitively insignificant need not bother with (I) or (II). But one need not subscribe to strong realist views to adopt (I) and (II). In particular, one need not accept the view that science does succeed in delivering true theories of the world to accept (I) and (II). Do bear in mind that (I) and (II) are semantic and not epistemological theses – though they can certainly help the realist epistemological cause. Be that as it may, my point of view will be a realist one. For I take very seriously Putnam's (1962) 'short argument' for theoretical terms, namely that the reason that theoretical terms are necessary is precisely that scientists employ terms like 'electron', 'virus', 'space-time curvature', and so on – and advance relevant theories – because they wish to *talk about* electrons, viruses, the curvature of space-time, and so on; that is, they want to find out about the deep structure of the world. It is the theoretical terms that provide scientists with the necessary linguistic tools for talking about things they *want* to talk about.

A tension, however, appears between (I) and (II). The standard descriptivist theories of reference (at least as they are assumed in the philosophy of science) satisfy (I) but not (II), while the standard causal theories of reference satisfy (II) but not (I).

The tragedy of the descriptivist theories of reference, when applied to scientific theories, is that while they honour the thought that theoretical terms do have factual reference (this being whatever satisfies certain theoretical descriptions), they face a genuine difficulty in honouring the thought that terms featuring in distinct theories have the same reference. But, I will argue, the culprit is *not* the idea that reference is fixed by descriptions – or better, that descriptions do play a significant role in reference-fixing. The culprit is *holism* – the view that the reference is fixed by the whole network of theoretical statements that a term is part of.

The tragedy of the causal theories of reference, when applied to scientific theories, is that while they honour the thought that the same (or different) terms that feature in different theories may well refer to the same worldly entity (this being the actual cause of whatever phenomena led to the introduction of a new theoretical entity and a term to refer to it), they face a genuine difficulty in explaining referential failure – insofar as there are causes of certain phenomena Φ, *any* term in *any* theory devised to stand for the cause of Φ does stand for these causes, irrespective of how wrong might be the theoretical descriptions associated with this term.

This is but an instance of a general failure of the causal theories to grant a genuine role to theories in reference-fixing. But, the culprit, I shall argue, is not the idea that causation is involved in the mechanism of reference-fixing. The culprit is the thought that the bare causal relation *in itself* is enough to fix the reference of theoretical terms.

After discussing the problems faced by the two standard theories of reference, as they are applied to the issue of the reference of theoretical terms, I will advance and defend causal descriptivism as an alternative account of the reference of the theory-dependent terms. I will argue that a hybrid theory of reference of theoretical terms (a) is independently well-motivated; and (b) meets in a satisfactory way conditions (I) and (II) above. Finally, I will unravel the key difficulties of a recent attempt to (dis)solve the issue of the reference of theoretical term by recourse to Ramsey sentences.

2 THE SWING OF THE PENDULUM I: DESCRIPTIVIST THEORIES OF REFERENCE

The story of the descriptivist theories of reference is well-known, so the reader can be spared a lot of details and subtleties. For present purposes, it is enough to recapitulate some of their key ideas. The characterization 'descriptivist *theories*' is not accidental, since the key ideas have undertaken considerable modifications over the years and in light of important philosophical controversies. The central thought, however, is that the competent speaker of a language (or the competent user of an expression) must know some identifying facts about the referent in order to refer successfully to it, where these identifying facts are captured by descriptions of the referent. If nothing satisfies the associated descriptions, then the expression does not refer to anything. But for an expression to refer to anything at all, it must be the case that its associated descriptions must be satisfied; that is, there must be something that has the properties attributed to it by the associated descriptions. It is useful to think of the associated descriptions as Fregean *modes of presentation*: the referent is presented as being in a certain way (and the same referent might be presented in different ways). We can even think of the modes of presentation as the *senses* of expressions; hence, senses are descriptive mechanisms that fix reference. That sense determines reference is then an important plank of descriptivist theories: an expression acquires its reference (if any) via its sense.

Until Kripke's (1972) well-known attack on the descriptivist theories, they were the only game in town. Though they were never used *explicitly* as theories of reference-fixing in the philosophy of science, they

were quietly operating in the background. For instance, the very (but short-lived) idea of providing explicit definitions of theoretical terms was but an application of the descriptivist theory of reference coupled with the idea that the associated descriptions should be couched in an observational vocabulary assumed to be independently meaningful and semantically kosher. Even more relaxed approaches to the meaning of theoretical terms – e.g. those based on Carnapian (1936) reduction sentences – were assuming the descriptivist theories. The key problem, however, that these looser approaches faced – namely that theoretical terms could not, in the end, be exhaustively defined by means of logical operations on observational statements – brought to light a shortcoming of the coupling of descriptivist theories with the allegedly sharp dichotomy between observational terms and theoretical ones. The problem was in the *coupling*, of course, since there is nothing in the descriptivist theories themselves that forbids the referent of an expression to be an unobservable entity. What matters is whether the associated descriptions are satisfied or not, not whether the satisfier can or cannot be described in a supposed privileged observational vocabulary. Indeed, the demise of the distinction between observational and theoretical terms (that is, the demise of the programme of providing distinct semantics for these allegedly distinct types of term) freed the descriptivist theories of a burden. The central thought that the burden of reference rests upon descriptive phrases was freed from the redundant and ill-motivated extra condition that these descriptive phrases should be couched in an observational vocabulary.

The way forward was to treat all vocabulary of scientific theories on a par and to claim that the descriptive phrases that fix the reference of theoretical terms are supplied by the theory as a whole: the meaning of a term is determined by its relations to other terms. More specifically, a seminal idea – due to Putnam – was that all theoretical concepts of a theory are 'law-cluster' concepts: they get their meaning via the plethora of nomological statements in which they occur. This view has two immediate consequences. The first is that since the nomological statements that constitute the 'law cluster' are synthetic, there is no way to separate them into two cleanly divided camps: those that fix the meaning of a concept and those that specify its empirical content. All do both. Hence, there is no way to draw the analytic–synthetic distinction *within* a theory. The second consequence is that the meaning (and hence the reference) of theoretical terms is fixed in a holistic way. The theoretical terms are implicitly defined, as it were, by the nomological statements that specify the

network of their connections within a theory – and, more specifically, by *all* of these nomological statements.

Semantic holism contributed significantly to the wide acceptance of the claim that theoretical discourse is meaningful and that theoretical terms have putative factual reference. But there are two shortcomings. The first has to do with the very idea of implicit definition. The second has to do with the extent of holism.

The chief attraction of implicit definitions is precisely that they fix meaning, not by analysing already known and understood concepts, but by legislating in a stipulative manner the truth of certain propositions, of which the defined concepts are constituents. Hence, they *create* or *constitute* meanings: for something to be an *F* (that is, for the concept of *F* to be applicable to it), such and such conditions must be satisfied. The chief drawback, however, is that an implicit definition is a kind of *indefinite description*: it defines a whole class of (or classes of) objects which can realize the formal structure, as defined by a set of axioms. There is no straightforward way in which unique satisfaction of the descriptions associated with the implicitly defined theoretical terms can be assumed. It might be thought that this latter problem is avoided at the point of application of the theory to the empirical world – at least this is what Schlick thought ([1932] 1979). But this is too quick. *First*, for it to work in the first place, the observational vocabulary (or a vocabulary fit to describe the empirical content of the theory) should be antecedently given and independent of the theory. If the meaning of this vocabulary too is fixed by the theory (that is, by the very same implicit definition that fixes the meanings of theoretical terms), then it can no longer anchor the theory to the empirical world. *Second*, even if the observational vocabulary is antecedently given and fixed independently of theory, it is still an open possibility that the theoretical structure that is fixed by the implicit definition is multiply realized.

We shall discuss these problems in more detail in Section 5. For the time being, let me focus on the issue of the extent of meaning holism. Here the otherwise liberating coupling of descriptivist theories with semantic holism becomes an unstable position. Descriptivist theories do warrant attributing reference to theoretical terms, but if semantic holism is rampant, the meaning of *all* terms (including those that normally count as observational) is specified in a holistic way. If semantic holism is moderate (if, that is, some terms get their meaning in a non-holistic way), a story needs to be told as to how this is possible which does not commit to semantic double standards. This story might well be possible.[1] Whether

or not semantic holism is rampant, the troublesome consequence that every time the theory changes, the meanings of *all* terms whose meaning is determined by the theory change too. We have then a thesis of radical meaning variance in theory change. Coupled with descriptivist theories of reference, and their key point that sense determines reference, semantic holism yields as a further consequence an even more radical thesis, namely reference variance.

The view that change of reference is *inevitable* when theories change – and hence the denial of the view that theoretical terms are transtheoretic – has been associated with Kuhn and Feyerabend. One of the standard examples concerns the terms 'mass', 'space', and 'time' as they occur in Newton's, and then in Einstein's, theory. As Kuhn ([1962] 1970, 102) famously stated:

> the physical referents of these Einsteinian concepts [i.e. mass, space, and time] are by no means identical with those of the Newtonian concepts that bear the same name.

Equally famously, Kuhn (1970, 128) went on to suggest that *observational terms* too, such as 'planet', have different meaning as well as reference when they occur in different theories (e.g. Ptolemy's and Copernicus').

This bizarre conclusion is, of course, the outcome of a dual commitment to semantic holism *and* the descriptivist theory of reference. Semantic holism entails that if there are changes in the theoretical/inferential network in which a term is embedded, the meaning of the term necessarily changes; the *new* meaning of the term is then given by its function within the *new* network.

Note that the descriptivist theories of reference are not *necessarily* holistic. Even when more than one description is associated with a term, it is allowed that individuals (the referents) are picked out by *weighted* descriptions. Then, not each and every change in the cluster of descriptions associated with the terms yields reference variance. Besides, descriptivist theories allow that two different descriptions may pick out the *same* individual (i.e. they may be coreferential) provided, of course, that they are not *inconsistent*.

So it is holism that does most of the work in getting radical reference variance. In particular, the culprits are the following two views:

(a) the network of nomological statements specify the descriptive sense of scientific terms in a *holistic* and undifferentiated manner; and
(b) changes in those networks are, always and necessarily, such that they yield *incompatible* descriptions of the same terms.

If these two theses are not independently supported, one cannot get radical reference variance. But there are good reasons to doubt both of them. Even Feyerabend (1965, 259) himself conceded that *not* all theoretical changes lead to changes in meaning and reference. He suggested that the rules (assumptions, postulates) of a theory form a hierarchy where more fundamental rules are presupposed by less fundamental ones. Then, only changes in *fundamental* rules lead to changes in meaning and reference. The fact is that a theory of radical reference variance is untenable *precisely* because it yields an implausible story about how science develops. Here are three reasons for the implausibility of this theory.

First, theories of radical reference variance entail that the referent of a term changes whenever there is even the *slightest* change in the network of descriptions in which the term is embedded. Hence, the theory leaves little, if any, room for *sameness* in reference. It entails that no current theoretical term can have the reference that in the past it was thought to possess, unless in the unlikely event it has retained *in full* its associated network of descriptions *and* nothing more has been added. Second, the theory entails that all scientific disputes about the features of a posited entity were mere equivocations. They could have been avoided had the relevant scientists been prudent enough to point out that they were talking about different entities! Third, it entails that there is no way to assess whether past scientists were right or wrong in their beliefs about the furniture of the world, unless their full web of beliefs has been retained – in which case, they were right!

This attempt to reduce the radical-reference-variance view to absurdity puts the blame on semantic holism and not on the descriptivist theories.[2] With this in mind, let us take a look at the rival causal account of reference and how it fares as a theory of reference-fixing for theoretical terms.

3 THE SWING OF THE PENDULUM II: CAUSAL THEORIES OF REFERENCE

In its original form, this was a theory of the reference of proper names. It was a *causal-historical* theory of proper names in the sense that it rendered the causal-historical chain of events that links current uses of a term to its introduction part of both the reference-fixing and the reference-transmission mechanisms. As introduced by Kripke, the causal-historical approach identifies the reference (denotation) of a name with the individual/entity

that this name was used – in an initial act of baptism – to denote. Hence, the reference of a currently used proper name, e.g. 'Aristotle', is fixed by the causal-historical chain which links the current use of the name with an introducing event, during which the name was given to its bearer (*Aristotle*). Descriptions which are, as a matter of fact, associated with the name (e.g. that Aristotle was the founder of the Lyceum) might be false, and yet current name users do refer to the named individual, insofar as their use of the name is part of a transmission chain which goes back to the introducing event. The thrust of the causal theory is that the relation between a word and an object is *direct* – a direct causal link – unmediated by a concept. In particular, the causal theory dispenses with descriptive senses as reference-fixing devices.

This theory is causal in an oblique sense, namely that the reference-transmission mechanism *is* causal. The reference-fixing mechanism is causal too but only in the sense that the introduction of the name is done (and hence its reference is fixed) in the *presence* – that is, the *perceived* presence – of its bearer. It's not as if the presence of a certain child *causes* its name-giver to give him the name 'Aristotle' as opposed to anything else. This part is purely (or almost purely, given several traditions of naming in Greece) conventional. Rather, what happens is that this conventional act of naming picks out its bearer uniquely because there is *causal contact* between the name given and the bearer of the name. So the key thought behind the causal theory of reference is that this causal relation – that underpins the causal contact between the name-giver and the dubbed item – fixes the reference in an unconceptualized way.

When Putnam extended this view to cover the reference of natural-kind and physical-magnitude terms, the idea was that the relevant introducing event (e.g. of the kind term 'gold') involved causal contact with samples of the dubbed stuff. But this is not clearly enough to fix the reference of a natural-kind term, the reason being that, at most, it succeeds in making the term refer to the sample of stuff that is actually present. To succeed in fixing the reference of the term to the natural kind itself (assuming that it is a natural kind) it is also required that in the very act of baptism the introduced term is said to refer to whatever stuff is similar to the one that causally grounds the introduction. So something like the following is needed during the introducing event:

> for all *x*, *x* is an *F* (e.g. an elephant) iff *x* stands in a specific similarity relation (e.g. sameness of nature) to *this* specific object, picked out by ostension.

As Putnam (1983, 73) put it, 'a term refers (to the object named) if it stands in the right relation (causal continuity in the case of proper names; sameness of "nature" in the case of kinds terms) to these existentially given things'.

Fixing the reference of physical-magnitude terms (e.g. 'electricity') is a variation of the same theme: when confronted with some observable phenomena, it is reasonable to assume that there is a physical magnitude (or entity) which causes them. Then we (or indeed, the first person to notice them) dub this magnitude with a term *t* and associate this magnitude with the causal production of these phenomena. During this introducing event of the term *t* as referring to this magnitude, one might typically associate *t* with a description, i.e. with a causal *story*, of the nature of the posited entity and of the properties in virtue of which it causes its paradigmatic observable effects. This initial description will most likely be incomplete, or even misguided. It may even be wrong: a mistaken account of the nature of this causal agent. But, on the causal theory, one has nonetheless introduced existentially a referent – an entity causally responsible for certain effects to which the term *t* refers. That is, one has asserted that:

> There is a ϕ [ϕ is causally responsible for certain effects Φ and for all terms *t* (*t* refers to ϕ if and only if *t* picks out the causal agent of Φ)].

The chief attraction of the causal theory is that it disposes of semantic incommensurability.[3] If, for instance, the referent of the term 'electricity' is fixed existentially, all different theories of electricity, refer to, and dispute over, the same 'existentially given' magnitude, namely the causal agent of salient electrical phenomena. The causal theory lends credence to the claim that even though past scientists had partially or fully incorrect beliefs about the properties of a causal agent, their investigations were continuous with those of subsequent scientists, since their common aim has been to identify the nature of the same causal agent.

The chief problem with the causal theories of reference is that they make referential success all too easy. If the reference of theoretical terms is fixed purely existentially, then insofar as *there is* a causal agent behind the relevant phenomena, the term is bound to end up referring to it. Hence, there can be no referential failure – even in cases where it is counter-intuitive to expect successful reference. Taken to its letter, the causal theory makes referential success inevitable. For instance, there is no easy way to show within a causal theory that 'phlogiston' was not referring to *oxygen*.

The problem is not with the idea that causation is involved in reference-fixing. It should clearly be part of the story of the referential failure of 'phlogiston' that phlogiston did *not* cause the phenomena for which the term 'phlogiston' was introduced. The problem is that the presence or absence of a cause – in and of itself – cannot be the total determinant of reference. Modes of presentation of the referent should be added.

Actually, the causal theory is consistent with the fact that descriptions may be used to identify a referent. Indeed, there are interesting cases in which the specification of a referent is (or can be) made *only* via a description, e.g. in the case of the introduction of the name 'Neptune' for the newly discovered planet that perturbed the orbit of Uranus, or of the name 'Jack the Ripper' (see Kripke 1972, 79–80, 96). Such cases, it should be noted, involve most occasions in which a theoretical entity is posited as the causal agent of certain phenomena and a name is picked to refer to it. It makes little difference that, as Kripke (1972, 106) put it, the description 'fixes the reference by some contingent marks of the object'. The important thing is that there are clear cases in which not much referential headway can be made without using descriptions.

4 THE PENDULUM'S RESTING PLACE: CAUSAL DESCRIPTIVISM

If we want to avoid the problems noted above, it seems we have to combine some lessons learned from the problems faced by the standard theories of reference. This would push us towards a causal descriptivist account of reference, one that utilizes insights and resources from causal and descriptivist approaches to reference-fixing.

4.1 The basics

Causal descriptivist accounts are by no means new. The basic *negative* idea behind them is this. The reference-fixing mechanism should not be just a non-conceptual causal relation to the referent; if it were, referential success would be easy to get (and trivialized) insofar as there is indeed some kind of causal contact with the referent. However, the reference-fixing mechanism should not be just a conceptual description-based relation to the referent; if it were, referential stability would be hard to get insofar as the description-based relation to the referent is determined by theories and theories change.

The basic *positive* idea, then, is this. The reference-fixing mechanism should have the following form:

$$R(x) = x \text{ causes phenomena } \Phi \textit{ and } D(x).^{4}$$

Term t refers to x if and only if $R(x)$.

This appears to be a genuine hybrid account. The causal relation with the referent plays an indispensable role in reference-fixing. This is meant to be captured by the claim that ⌜x causes Φ⌝. But the causal relation is not enough. Reference is fixed by means of descriptions of the causal role attributed to the putative referent. This is meant to be captured by $D(x)$. $D(x)$ is a causal description, namely a description of the ways in which the posited referent is supposed to be causally connected with the phenomena Φ that it is taken to cause.

That *each* component is required may be defended as follows. The causal component ⌜x causes Φ⌝ is required because the referents of theoretical terms are introduced as causes of certain phenomena. The descriptive component $D(x)$ is required because the referent (the cause of Φ) should be attributed some properties – those that capture its causal role – if cognitive (as opposed to merely causal) access to it is to be had. That *both* components are required may be seen from the fact that their verdicts may not match: the entity that satisfies $D(x)$ vis-à-vis Φ might not be the entity that does cause Φ. The oxygen/phlogiston case is quite instructive in this respect. It is quite clear that in this case, there is a mismatch between the two components of $R(x)$, if they are taken separately: the cause x of the phenomena Φ of combustion was the presence of oxygen and yet the (phlogiston-based) description $D(x)$ of the cause could not single out oxygen as the cause. The presence of both components of $R(x)$ makes sure of the following: A term t refers to x only if x is the cause of Φ. If causal description $D(x)$ is satisfied by some entity x, but x is not the cause of Φ, then t does not refer to it. Conversely, if something does indeed cause Φ but $D(x)$ is not satisfied by this something, then t does not refer to it.

4.2 Three objections

Before I elaborate further on the basic idea, let me forestall three plausible objections to it. The first is that this hybrid theory is ad hoc, its only motivation being to accommodate counterexamples to the two standard theories of reference. The reply to this objection is that it seems a plausible general condition on reference-fixing that it combine two elements that need not coincide in all cases, namely a causal element and a

cognitive one.[5] In ordinary cases of perceptual access to objects – which are the paradigm cases of reference-fixing – the two conditions coincide (or tend to coincide anyway). Perception is a process by means of which we are in causal contact with worldly objects and by means of which we get cognitive access to them and their properties. Even if we do not buy into causal theories of knowledge, it is generally true that the perceptual processes by means of which we get to know things about objects of perception are the very processes by means of which we are causally connected with the objects of perception. There are malign cases, but these need not invalidate the general point. In the case of perceptual contact, R(*x*) is independently plausible and, as a rule, correct. In the case of fixing the reference of theoretical terms, these two conditions diverge. This is simply because there is no perceptual contact with the referent, though there is causal contact with it (if indeed, there is one). Hence, there must be a way in which cognitive access to the referent is ensured, at least in principle. Otherwise, the entity that does the causing of Φ might not be, even in principle, cognitively accessible. Then, referent-fixing would be a merely existential act: there is something that causes Φ. The cognitive access to this something – and hence the very possibility of knowing what this something is and whether it is indeed the one required by the truth of the theory which posits it – would require some kind of description of it: a description which would offer some *identifying marks*, those, I would add, by means of which it is supposed to play the causal role it does. R(*x*) then becomes independently plausible: it ensures the cohabitation of the two elements of reference-fixing, i.e. the causal and the cognitive.

The second plausible objection is that the account offered is not *really* a hybrid one – it is, in essence, a *descriptive account* of reference-fixing. The reference, it might be said, is fixed by a causal *description*. A variant of this objection, voiced by Raatikainen (2007) is something we already considered, namely that the original causal theory of reference did allow for descriptions to play a role in reference-fixing but insisted that descriptions may well be insufficient and unnecessary for a unique identification of the referent. On reflection, I think nothing much hangs on this objection. If the causal element in the original causal theory of reference is reduced to the role that causation plays in the *transmission* of reference of a term from those present in the act of baptism to its current users, as Raatikainen implies, this is a very thin causal theory of reference. If, as is more plausible, causation is said to play a more central role in reference-fixing, by linking the name with its referent without the need to interpolate an identifying description, then the objections noted above do suggest

that there is more to reference-fixing than causal contact. In any event, if descriptions are allowed to play some role in reference-fixing, it's a semantic issue whether the resulting theory is *really* descriptive.[6] The way causal descriptivism was articulated above does give some genuine role to causation, even if, at the end of the day, it is captured by claims of the form '*x* causes Φ'.

The third plausible objection is certainly the most serious and may be called 'the too little/too much' objection. In essence it is the problem of how rich the description should be, given that it is not the *whole* of the theory (as holistic accounts implausibly required). If the causal description D(*x*) associated with a newly posited entity is very rich, even though it may explain in virtue of what putative mechanisms it is supposed to play its causal role, it will be difficult to ensure that it is satisfied by anything, let alone to guarantee some referential continuity in theory change. The entities posited by successor theories rarely ever take up most of the explanatory-causal structure attributed to abandoned entities. If, on the other hand, the causal description attributed to a newly posited entity is too thin, it is quite likely that it will be multiply satisfied. Hence, there will be no guarantee that a unique referent is picked out.

The obvious answer to this objection is this: D(*x*) should have the right size! It should be neither too rich, not too thin. But then how is this possible? It is possible because descriptions associated with theoretical terms play a certain *role* in reference-fixing: they are meant to offer enough identifying markers of an entity (related to its causal role vis-à-vis phenomena Φ) to allow the stable use of the term in certain inductive and explanatory practices; but they are not meant to asphyxiate the putative referent, that is, to leave no room for error, ignorance, or improvement. Theory development is an ongoing process and, I claim, it is an ongoing concern of the scientists to maximize the chances of referential continuity of theoretical terms (without, however, making referential continuity inevitable). Hence, care is generally taken to associate with a term enough descriptions to pick out its referent uniquely and to ground its putative explanatory role, but not too many to forestall any referential continuity. This is evidenced by the fact that, as a rule, there is consensus vis-à-vis referential success and failure among scientists, which means that there is consensus as to which were the reference-fixing descriptions and which were not. As a quick example of this, take the much-discussed contrasting cases of 'caloric' and 'electricity'. They were both described as imponderable fluids;

they were both taken to have material composition; etc. But it turned out that though these descriptions were reference-fixing for 'caloric', they were not reference-fixing for 'electricity'. When nothing turned out to satisfy them in the case of the cause of heat phenomena, 'caloric' was abandoned as denotationless. But 'electricity' was not abandoned as denotationless, even though the foregoing descriptions were not satisfied; hence, they were *not* reference-fixing descriptions.

Admittedly, there is an element of hindsight in this way of putting things. But it is hindsight available to contemporary practitioners themselves. Here are two dictionary entries of the terms 'caloric' and 'electricity' in 1832:

> CALORIC, in Chemistry, a modern term introduced into philosophy, to denote that substance, by the influence of which are produced all the phenomena of heat, and which was formerly distinguished by the term igneous fluid, matter of heat, and other analogous demonstrations. (Jamieson 1832, 140)

> ELECTRICITY, the name of an unknown natural power which produces a great variety of peculiar and surprising phenomena, the first of which are supposed to have been observed in the mineral substance called amber, whence they have been denominated electrical phenomena, and the laws, hypotheses, experiments, &c by which they are explained and illustrated, form together the science of electricity. (Jamieson 1832, 255)

Notice the important difference in the way the two referents are described. In the case of 'caloric', it is part of the identifying markers of it that it is a (material) substance. In the case of 'electricity' it is not. Rather, electricity is picked out by descriptions that allow refinement of the putative nature of the power that produces certain phenomena.[7]

More generally, we can follow David Papineau's (1996) suggestion that the descriptions associated with a theoretical term are divided into three parts: the 'yes-part', that is, those that do contribute to the definition of the term; the 'no-part', that is, those that definitely do not contribute to the definition of the term; and, finally, the 'perhaps-part', that is, those descriptions which might (or might not) contribute the definition of the term. Insofar as the yes-part is rich enough to pick out a referent uniquely, the presence of the perhaps-part (which yields a certain imprecision in the meaning of the term) does not endanger the referential determinacy of the term.[8] What the above discussion has added is that the determination of the yes-part of the definition of a term is by and large brought about by considerations of maximizing the chances of transtheoretic referential stability without making referential success inevitable.

4.3 Some meat on the bones

I take it that an important feature of the causal descriptivist account outlined above is the tracking requirement: a theoretical term *t* must track its referent.[9] This means that it must allow the determination of the reference, the acquisition of further information about it, its re-identification, etc. It is obvious that the satisfaction of this requirement presupposes the use of identifying descriptions.

In my (1999) *Scientific Realism*, I developed causal descriptivism on the basis of Enç's (1976) claim that the burden of reference is borne by the *kind-constitutive properties* attributed to the posited entity. Hence, I claimed that the causal descriptions that fix the reference of a term – what I now call D(*x*) – should capture these properties. This way to proceed can explain referential success and failure. 'Phlogiston' fails to refer because no entity has the kind-constitutive properties attributed to phlogiston. And 'oxygen' succeeds in referring, because there is an entity with oxygen kind-constitutive properties. Since we have no theory-independent access to the kind-constitutive properties of a natural kind, we have to rely on theories and their causal-explanatory descriptions of the entities they posit. There is simply no other way to proceed. But this does not imply that the whole of the theory is implicated in reference-fixing. The kind-constitutive properties are those whose presence makes a set of objects have the same, or sufficiently similar, manifest properties, causal behaviour, and causal powers. Their specification need not involve the whole of the theory.

If we assume that natural kinds have boundaries – based on objective similarities and differences – we can see how causal descriptions succeed in fixing the reference of a term. The one and only entity to which the term refers is the entity which is characterized by the relevant kind-constitutive properties. So, D(*x*) identifies the referent *x* of term *t* in such a way that (a) if no entity satisfies D(*x*), (i.e. if it is true of no entity), then *t* does not refer; and (b) if an entity *y* does not satisfy D(*x*), *y* and *x* cannot play the same causal role. So, on this view, we have a readily available account of referential success and failure.

I still think this idea is essentially correct. But I want to broaden it a bit by claiming that the burden of reference is carried by *any* stable identifying properties provided that they are taken to contribute to the causal role attributed to the posited entity vis-à-vis the phenomena Φ. This loosening is important because what properties are kind constitutive might be a matter that is settled at quite advanced stages of inquiry whereas successful reference can be established at quite early stages. It is hoped

that the identifying properties will be part of the kind-constitutive properties of the posited entities, but this is not something that has either to be assumed *ab initio* or required in the end. Actually, this kind of latitude allows for reference refinement.

There are two prima facie plausible objections to the claim that stable contingent properties may play a role in reference-fixing. Relying on Kripke's well-known modal argument against descriptivist theories, one may argue that in modal contexts where the referent does not possess these contingent properties, reference may still succeed, given the causal connection. Hence, it might be claimed, there is a danger that the suggested account is descriptivist only in name. Similarly, relying on Kripke's well-known epistemic argument against descriptivist theories, one may argue as follows: if we are wrong about the descriptions which concern putative contingent properties, but still in causal contact with the referent, do we succeed in referring to it? If yes, doesn't this show that it is the causal element that is essential to reference? And, if not, why not?

Though both objections deserve more attention than I offer here, I think meeting them should start by stressing that in the suggested account causation and stable identifying properties (picked out by descriptions) work *in tandem* in reference-fixing. Regarding the first objection, it is precisely an advantage of this account that reference is not fixed by descriptions only. Causal contact serves the purpose of anchoring the descriptions to the referent, thereby rendering these descriptions (at least in the favourable cases) descriptions *of* the referent in the actual world. Conversely (and regarding the second objection), it should be clear that referential success is not merely a matter of causal contact since, as stressed already, it requires that the theoretical terms track their referents, which is achieved by securing identifying marks of the referents, however contingent. To put it succinctly, causal contact anchors the descriptions to the referent; descriptions enable tracking the referent.

To illustrate all this let's discuss briefly the neglected case of chlorine. The term 'chlorine' was first introduced by Humphry Davy in 1812 to refer to what was taken to be dephlogisticated muriatic acid. Carl Wilhelm Scheele had isolated choline already in 1774, but being based on the phlogiston theory, he described it as dephlogisticated muriatic acid. He produced it by heating manganese dioxide with hydrochloric acid (called 'muriatic acid'). The idea was that the heating had removed phlogiston from the muriatic acid. There was some controversy as to whether dephlogisticated muriatic acid was an element or a compound. Davy

experimented with dephlogisticated muriatic acid, became convinced that it was an element and called it 'chlorine'. Actually, the very term 'chlorine' was introduced to capture the greenish-yellowish colour of the gas (from the Greek word 'chloros', meaning 'green'). But Davy was fully aware of two things. One: the gas he had isolated was the same one that Scheele had isolated.[10] Two: Scheele's description of it (and the general theory on which it was based) was flawed; the key problem was that dephlogisticated muriatic acid (aka *oxymuriatic acid*) did not contain oxygen. It was an elementary substance.[11] Hence, Davy thought that a change of name was appropriate too.[12] In particular, the change of name was meant to suggest not a change of the reference, but a change of how the referent was identified via appropriate properties of it. The detailed story is long and intricate, as it involved a number of experiments in which Davy showed that chlorine did not contain oxygen and that neither did hydrochloric acid. But the bottom line is this. Chlorine is an element (and not a compound); it does not involve oxygen; it has bleaching power (when water is present); it is a constituent of sea salt; it unites with hydrogen (in equal volumes) to produce hydrochloric acid; etc. This looks very much like an associated description of the form D(x) above. It is a description of a causal agent of a certain sort. It is a description by means of associated properties and relations. Davy was open to the fact that these might prove wrong. He was almost certain that they would need refinement, anyway. Ultimately, the reference should be fixed by the kind-constitutive properties of chlorine.[13] But until this is done, there is no reason to think that reference has not been fixed by means of stable characteristic properties. Indeed, it is clear that all subsequent discussion relied on the fixity of the reference of 'chlorine' to an element and not a compound.[14] As is well-known, the atomic weight of chlorine, being 35.5 presented an anomaly to Proust's law of definite proportions. It turned out that samples of chlorine contained two stable chlorine isotopes (chlorine-35 and chlorine-37 in a ratio of 3:1).

This development is instructive when it comes to reference refinement. Isotopes are atoms with the same number of protons but different numbers of neutrons in their nucleus. Isotopes of the same element have the same chemical properties, but different physical ones, e.g. atomic mass. The very term 'isotope' was introduced to refer to entities (atoms) with different causal roles vis-à-vis various radioactive phenomena but 'located' in the same place in the periodic table, that is having the same atomic number and chemical properties. It turns out that two kinds of chlorine play the causal role specified by the description D(x) noted above: they were different in virtue of properties that differentiated them qua *nuclides*.

A nuclide is an atom with a specific number of protons and neutrons in its nucleus. Hence, there are two chlorine nuclides and it is arguable that they capture a finer division into kinds of chlorine. So the initial term 'chlorine' admits of reference refinement which is achieved when the original identifying properties are refined in such a way that there is a better match between them and the kind-constitutive properties.

I take it that the foregoing account has a number of genuine advantages.

- It does not make referential continuity inevitable.
- It explains both referential failure and success in a non-trivial way.
- It allows for a further determination of the referent by means of added-on descriptions.
- It leaves open the possibility of re-identifying the reference of a term, e.g. when it turns out that $x = y$ even though $D(x)$ is different from $D'(y)$. This is likely to happen when there is unifying descriptivist that brings under its fold both $D(x)$ and $D'(y)$.
- It shows how the associated descriptions $D(x)$ track the referent of the term: this happens when they do succeed in identifying the causal agent of phenomena Φ.

5 RAMSEY-SENTENCE REALISM AND REFERENCE

There is a view of growing popularity that a Ramsey-sentence approach to theories will (dis)solve the issue of the reference of theoretical terms. Actually, this approach is taken to be part of a broader descriptivist account of reference. Since the readers might not be familiar with the Ramsey-sentence approach, here is a brief summary of it.

To get the Ramsey sentence ^{R}TC of a (finitely axiomatizable) theory TC we conjoin the axioms of TC in a single sentence, replace all theoretical predicates with distinct variables u_i, and then bind these variables by placing an equal number of existential quantifiers $\exists u_i$ in front of the resulting formula. Suppose that the theory TC is represented as TC (t_1, …, t_n; o_1, …, o_m), where TC is a purely logical m+n-predicate. The Ramsey sentence ^{R}TC of TC is: $\exists u_1$ ($\exists u_2$ … ($\exists u_n$TC (u_1, …, u_n; o_1, …, o_m). For simplicity let us say that the theoretical terms of TC form an n-tuple $t = \langle t_1, \ldots, t_n \rangle$, and the o-terms of TC form an m-tuple $o = \langle o_1, \ldots, o_m \rangle$. Then, ^{R}TC takes the more convenient form: $\exists u$TC(u,o).

Now, as Carnap showed, the theory in the old form (i.e. TC) is *logically equivalent* to the conjunction ^{R}TC & (^{R}TC→TC). ^{R}TC is the Ramsey

sentence of the theory, while the conditional $^{R}TC \rightarrow TC$ – known as the Carnap sentence – asserts that *if* there is a class of entities that satisfy the Ramsey sentence, *then* the theoretical terms of the theory denote the members of this class. Carnap suggested that the Ramsey sentence of the theory captured its factual content (which was expressed in the rich language LT), while the conditional $^{R}TC \rightarrow TC$ captured its analytic content (it is a *meaning postulate*).[15]

Modulo the Carnap sentence, the theory is equivalent to its Ramsey sentence. But no theoretical terms occur in the Ramsey sentence of the theory (by construction). Hence, it might be thought, the problem of the reference of theoretical terms is (dis)solved. Actually, there are two ways to proceed, given the foregoing equivalence. The first is eliminative, the second is vindicatory. According to the eliminative way, the issue of the reference of theoretical terms vanishes. According to the vindicatory way, the issue of the reference of theoretical terms is solved once and for all. Let us take them in reverse order.

Given the equivalence TC iff ^{R}TC & ($^{R}TC \rightarrow TC$), we can think of the theory as implicitly defining its theoretical terms. An implicit definition of a concept F (or a set of concepts F, G, H, …) fixes the meaning of this concept specifying that certain postulates in which it occurs are true (similarly, for a set of concepts). Using standard notation, we may write '#_' for the set of postulates with a blank wherever the *definiendum*, F, occurs; the implicit definition then assumes the form '$F =_{def} \#F$ is true'. This kind of move implies that the entities to which the implicitly defined concepts apply are whatever ones satisfy the postulates (provided, of course, that the postulates are consistent). Actually, *any* system of entities that satisfy the postulates is such that the implicitly defined concepts apply to them. The Ramsey sentence ^{R}TC says that there are classes of entities which are correlated with the observable events in the way the postulates of the theory describe; but it does not say what exactly those entities are – it does not pick out any such class in particular. The Carnap sentence, then, asserts that '*if* anything satisfies the Ramsey sentence of the theory, *then* the entities posited by the theory do'. Hence, the theoretical concepts of the theory are implicitly defined by the Carnap sentence in such a way that they refer to *whatever* entities satisfy the Ramsey sentence. This is clearly an implicit *definition*: it stipulates meanings without defining them explicitly.

It seems the only residual question is this. Are there entities that satisfy the Ramsey sentence of the theory? David Lewis (1970), who made the above approach popular, urged that theoretical terms refer only if there is

a *unique* realization of TC, that is, only if there are no multiple classes of entities which stand in relation TC to *o*. If there are more than one such classes, the theoretical terms should be taken to be denotationless. His motivation for this claim was that such a view is the lesser of two evils. In case of non-unique realization, there is no non-arbitrary way to pick one realization. So, Lewis thought, we are forced to accept either that theoretical terms do not name anything, or that they name the elements of one arbitrarily chosen realization. For him, however, 'either of these alternatives concedes too much to the instrumentalist view of a theory as a mere formal abacus' (1970, 432).

If all goes well – if, that is, *uniqueness is achieved* – theoretical terms are implicitly defined *and* their reference is secured: to the unique realization of the theory. That's why I called this kind of view 'vindicatory'. All does *not* go well, however. Since the reason why all does not go well is related to the prospects of the eliminative approach, I should look at it first.

On the eliminative view, the very idea that the Ramsey sentence *replaces* the theory shows that the very issue of the referential success or failure of theoretical terms becomes irrelevant to the appraisal of theory and its claims to truth. All that is required for the truth of the theory is that the Ramsey sentence is satisfied: that there is a realization of it. However, as Lewis was quick to note, the Ramsey-sentence might not have an *exact* realization; it might have a near-realization. Let ^{R}TC be the Ramsey sentence of the original theory which has no unique realization, and TC* a theory which is a modification of TC (perhaps, a corrected version of TC). Suppose that TC* has a unique realization R which happens to nearly realize TC. Then, R is a near-realization of ^{R}TC. In the shift from TC to TC* there is, typically, change of reference (since the realizations of TC and TC* are different) unless the realization of TC* is a near-realization of TC too. In this case, it can be plausibly argued that there is referential continuity between TC and TC*. Here is exactly where the eliminative view gets its grip. If the content of the theory is *fully* captured by its Ramsey sentence, then – it is argued – referential failure (or success, for that matter) becomes irrelevant: the realizer of ^{R}TC* is the near-realizer of ^{R}TC and *that's it*. The issue of reference drops out of the picture. Pierre Cruse (2004, 140), who – together with David Papineau (see Cruse and Papineau 2002) – defends this view, goes on to claim that there needn't even be a near-realizer; various realizers might be allowed and the terms of the theory might be taken to 'partially denote' each of them.

The key difference between the vindicatory way and the eliminative way concerns the status of the Carnap sentence ^{R}TC→TC. The eliminative

approach takes it to be 'devoid of cognitive content' (Cruse 2004, 146). The vindicatory approach takes it to be cognitively significant. In fact, in Lewis' hands, the Carnap sentence is strengthened by the uniqueness-of-realization requirement. It now reads as follows: '*if* anything uniquely satisfies the Ramsey-sentence of the theory, *then* the entities posited by the theory do'. Let's call this the Lewis sentence.[16] Accordingly, the theoretical terms are fully defined (in fact, they can be *explicitly* defined) and their reference is fixed. The original Carnap sentence does not imply uniqueness of realization. It is fully consistent with it that there might be multiple realizers of the theory TC.[17] If there happen to be multiple realizers, the vindicatory approach says the terms of TC fail to refer. If there are multiple realizers, the Carnap sentence does allow that the terms of the theory might refer to them. In fact, as noted above, the Carnap sentence has no means to single out one of the many realizations as the 'correct' one.

The real issue, however, is not whether the Carnap sentence has cognitive content or not – though it does: the Carnap sentence poses a certain restriction on the class of models that satisfy the theory; it excludes from it all models in which the Carnap sentence fails. The real issue is with *multiple realization*. It turns out that the Ramsey sentence of the theory is bound to be multiply realized, unless restrictions are imposed on the range of its (second-order) variables. Suppose the Ramsey sentence ^{R}TC of a theory TC is empirically adequate. Subject to certain cardinality constraints, if the Ramsey sentence is empirically adequate, it is true. In other words, there is at least one realisation of it and, given plausible assumptions about the universe of discourse of the second-order variables, a multiplicity of them. The proof of this has been given in different versions by several people.[18] Its thrust is this. Since ^{R}TC is consistent, it has a model. Call it M. Take W to be the 'intended' model of TC and assume that the cardinality of M is equal to the cardinality of W. Since ^{R}TC is empirically adequate, the observational submodel of M will be identical to the observational submodel of W. That is, both the theory TC and its Ramsey sentence ^{R}TC will 'save the (same) phenomena'. Now, since M and T have the same cardinality, we can construct a one-to-one correspondence f between the domains of M and W and *define* relations R' in W such that for any theoretical relation R in M, $R(x_1 \ldots x_n)$ iff $R'(fx_1 \ldots fx_n)$. We have induced a structure-preserving mapping of M on to W; hence, M and W are isomorphic and W becomes a model of ^{R}TC.[19]

The Ramsey sentence of a theory may *fail* to be empirically adequate. No one denies that the Ramsey sentence of the theory has empirical content or that it might be empirically false. But the point still remains that if

it *is* empirically adequate (if, that is, the structure of observable phenomena is embedded in one of its models), then it is bound to be true. For, as we have seen, given some cardinality constraints, it is guaranteed that there is an interpretation of the variables of ^{R}TC in the theory's intended domain.

This result has a straightforward bearing on both the eliminative and the vindicatory approaches. I take it that it simply undermines the eliminative approach. For it is not just the issue of reference that drops out of the picture. The issue of realism goes by the board too: empirical adequacy collapses to truth. And not just that. There is simply no way to tell even what kind of things the second-order variables of the Ramsey sentence range over. When it comes to the vindicatory approach, things are somewhat more complicated. It is in the spirit of this approach that there should be ways to exclude various interpretations of the Ramsey-sentence variables, aiming to constrain them, at least in principle, to just one. The way to do this, as Lewis emphasized in subsequent work, is to impose causal constraints on the admissible interpretations of the language. That is, it is necessary to move away from purely descriptivist accounts of reference-fixing.

As Lewis (1984) has stressed, the causal constraint on an admissible referential scheme is not just more theory – a causal theory – added to the theory of the world. Rather, it is constraint to which a referential scheme (a realization of the theoretical terms of the theory) must conform, if it is to be the correct one. Perhaps the best way to understand what this constraint might be has been suggested by G. H. Merrill (1980), who argued that it is questionable that realism conceives of the world merely as *a set of individuals*, i.e. as a model-theoretic *universe of discourse*. The world, a realist would say, is a *structured entity*. Its individuals stand in specific relations to one another, or to subsets of individuals. In particular, a realist would assert two things: (1) the constituents of the world (individuals and properties/relations) are independent of any particular representation we have of it; and (2) the world is *already* structured, *independently* of the language. Under these assumptions, it is easy to see how the Ramsey sentence of the theory might fail to be true, even though it might be empirically adequate. Among all those relations-in-extension which satisfy the Ramsey sentence, only those which express *real relations* should be considered. But specifying which relations are real requires knowing something *beyond* the Ramsey sentence of the theory, namely which extensions are 'natural', i.e. which subsets of the power set of the domain of discourse correspond to natural properties and relations.

The view that the world is a *structured domain* places a *constraint* on the interpretation of a language, namely that referents must be *eligible*. Lewis' eligibility constraint is *prior* to the interpretation of a language and such that an interpretation must satisfy it in order to be intended. Lewis does not suggest an eligibility *theory* which is open to reinterpretations. He offers a constraint and suggests that we have to turn to physics and its 'inegalitarianism' in order to find the elite things and classes of things that constitute the joints of the world. Then he argues that an interpretation is unintended – and it would be disqualified – if it employs gerrymandered referents, i.e. putative referents that do not belong to the objective structure of the world. It follows that if 'we limit ourselves to the eligible interpretations, the ones that respect the objective joints in nature, there is no longer any guarantee that (almost) any world can satisfy (almost) any theory' (1984, 227). There is now a way to eliminate many of the possible realizers of the Ramsey sentence: only those that are eligible remain, where the eligibility of a realizer is a function of the eligibility of the classes over which the variables of the Ramsey sentence vary.

6 CONCLUSION

From a realist point of view, an important constraint on the theory of reference – if not the *raison d'être* of a theory of reference – is the accommodation of linguistic practices to the causal structure of the world. Given that this accommodation is, as a matter of fact, an historical process, it is vital for a realist view of science that the theory of reference allows for convergence in the scientific image of the world. Theoretical terms are necessary for scientists to talk about posits not given in experience; but they should be taken to have transtheoretical robustness. Unless continuity in theory change and convergence are established, past failures of scientific theories will act as defeaters of the view that current science is on the right track. The causal descriptivist approach to the reference of theoretical terms is fit to strike a balance between, on the one hand, the theory dependence of theoretical terms and, on the other hand, viewing theoretical terms as relevantly transtheoretical. This is reason enough to be worthy of development.

NOTES

[1] The best teller of this story is Fodor 1984.

[2] It might be thought that the description theories are not good theories of reference anyway. For an effective rebuttal of this claim, see Jackson 1998.

[3] For definitive work on the issue of incommensurability see Howard Sankey's papers, e.g. 2009.

[4] A variant of this account is considered but not adopted by Kroon (1985).

[5] This is also suggested by Kroon (1985, 1987, 2009).

[6] Interestingly, the mirror image of this objection is made by Boyd (2010, 224), who claims that causal descriptivist theories are really *causal* theories, since descriptions and their use play a causal role in reference-fixing. Deploying descriptions may well be part of the causal profile of reference-fixing by language users. But the key point is that descriptions offer identifying marks without which the referent cannot be tracked.

[7] This is by no means atypical. Here are the relevant entries in Andrew Ure's *Dictionary* of 1820.

> CALORIC. The agent to which the phenomena of heat and combustion are ascribed. This is hypothetically regarded as a fluid, of inappreciable tenuity, whose particles are endowed with indefinite ido-repulsive powers, and which, by their distribution in various proportions among the particles of ponderable matter, modify cohesive attraction, giving birth to the three general forms of gaseous, liquid, and solid. ([1820] 1828, 251)

> ELECTRICTIY. The phenomena displayed by rubbing a piece of amber, constitute the first physical fact recorded in the history of science. … From electron, the Greek name of amber, has arisen the science of electricity, which investigates the attractions and repulsions, the emission of light, and explosions, which are produced, not only by the friction of vitreous, resinous, and metallic surfaces, but by the heating, cooling, evaporation, and mutual contact of a vast number of bodies. ([1820] 1828, 401)

Ure, very meaningfully, adds that in the case of caloric there is an alternative account of the causes of heat phenomena, i.e. the vibratory motion of the 'particles of common matter'.

[8] There are certain difficulties with Papineau's account which are discussed by Papineau himself and by Kroon and Nola (2001).

[9] This tracking requirement is central in Boyd's (2010) accommodationist view of reference of natural-kind terms.

[10] 'It is evident from this series of observations, that SCHEELE'S view (though obscured by terms derived from a vague and unfounded general theory) of the nature of the oxymuriatic and muriatic acids, may be considered as an expression of facts' (Davy 1902, 28).

[11] 'I stated a number of facts, which inclined me to believe, that the body improperly called in the modern nomenclature of chemistry, oxymuriatic acid gas, has not as yet been decompounded; but that it is a peculiar substance, elementary as far as our knowledge extends, and analogous in many of its properties to oxygene gas' (Davy 1902, 40).

[12] 'To call a body which is not known to contain oxygene, and which cannot contain muriatic acid, oxymuriatic acid, is contrary to the principles of that nomenclature in which it is adopted; and an alteration of it seems necessary to assist the progress of discussion, and to diffuse just ideas on the subject. If the great discoverer of this substance had signified it by any simple name, it

would have been proper to have recurred to it; but, dephlogisticated marine acid is a term which can hardly be adopted in the present advanced era of the science' (Davy 1902, 59).

[13] 'As chemistry improves, many other alterations will be necessary; and it is to be hoped that whenever they take place, they will be made independent of all speculative views, and that new names will be derived from some simple and invariable property, and that mere arbitrary designations will be employed, to signify the class to which compounds or simple bodies belong' (Davy 1902, 61–62).

[14] Cf. Turner 1835, 224–226.

[15] For more on this, see Psillos 2000, 2008, and Psillos and Christopoulou 2009.

[16] The Lewis sentence is implied by the Carnap sentence. Lewis adds two more meaning postulates. (1) *If* nothing satisfies the Ramsey sentence of the theory, *then* the entities posited by the theory do not satisfy it. Hence, the terms of the theory are denotationless. Postulate (1) is independent of the Carnap sentence. (2) *If* the Ramsey sentence of the theory is multiple realized, *then* the terms of the theory are denotationless. Postulate (2) is actually inconsistent with the Carnap sentence in that it disallows that the terms of the theory might refer to the entities posited by the theory, if the theory is not uniquely realized by these entities.

[17] Carnap famously thought that there will be mathematical realizers of the theory; see my 2008 for the details.

[18] Winnie 1967, 226–227; Demopoulos 2003, 387; Ketland 2004.

[19] This is a version of the so-called Newman problem that Russell's structuralism faced. For more on this, as well as on Ramsey sentences, see my 2006. Cruse's reaction to this claim is given in his 2005. For an effective rebuttal see Ainsworth 2009.

REFERENCES

Ainsworth, Peter (2009) Newman's objection. *British Journal for the Philosophy of Science* 60: 135–171.

Boyd, Richard (2010) Realism, natural kinds, and philosophic methods. H. Beebee and N. Sabbarton-Leary (eds.) *The Semantics and Metaphysics of Natural Kinds*. London: Routledge.

Carnap, Rudolf (1936) Testability and meaning. *Philosophy of Science* 3: 419–471.

Cruse, Pierre (2004) Scientific realism, Ramsey sentences and the reference of theoretical terms. *International Studies in the Philosophy of Science* 18: 133–149.

(2005) Ramsey-sentences, structural realism and trivial realisation. *Studies in History and Philosophy of Science* 36: 557–576.

Cruse, Pierre and Papineau, David (2002) Scientific realism without reference. In M. Marsonet (ed.) *The Problem of Realism*. Aldershot: Ashgate, pp. 174–189.

Davy, Humphry (1902) *The Elementary Nature of Chlorine: Papers by Humphry Davy (1809–1818)*. Edinburgh: Alembic Club.

Demopoulos, William (2003) On the rational reconstruction of our theoretical knowledge. *British Journal for the Philosophy of Science* 54: 371–403.
Enç, Berent (1976) Reference of theoretical terms. *Noûs* 10: 261–282.
Feyerabend, Paul (1965) Problems of empiricism. In R. Colodny (ed.) *Beyond the Edge of Certainty*. Englewood Cliffs, NJ: Prentice-Hall.
Fodor, Jerry (1984) Observation reconsidered. *Philosophy of Science* 51: 23–43.
Jackson, Frank (1998) Reference and description revisited. *Philosophical Perspectives* 12: 201–218.
Jamieson, Alexander (1832) *A Dictionary of Mechanical Science*. London: Henry Fisher, Son & Co.
Ketland, Jeff (2004) Empirical adequacy and Ramsification. *British Journal for the Philosophy of Science* 55: 287–300.
Kripke, Saul (1972) Naming and necessity. In D. Davidson and G. Harman (eds.) *Semantics of Natural Language*. Dordrecht: Reidel, 253–355.
Kroon, Fred (1985) Theoretical terms and the causal view of reference. *Australasian Journal of Philosophy* 63: 142–166.
(1987) Causal descriptivism. *Australasian Journal of Philosophy* 65: 1–17.
(2009) Names, plans, and descriptions. In D. Braddon-Mitchell and R. Nola (eds.) *Conceptual Analysis and Philosophical Naturalism*. Cambridge, MA: MIT Press, pp. 139–158.
Kroon, Fred and Nola, Robert (2001) Ramsification, reference fixing and incommensurability. In P. Hoyningen-Huene and H. Sankey (eds.) *Incommensurability and Related Matters*. Dordrecht: Kluwer, pp. 91–121.
Kuhn, T. S. ([1962] 1970) *The Structure of Scientific Revolutions*, 2nd enlarged edn. University of Chicago Press.
Lewis, D. (1970) How to define theoretical terms. *Journal of Philosophy* 67: 427–446.
Lewis, David (1984) Putnam's paradox. *Australasian Journal of Philosophy* 62: 221–236.
Merrill, G. H. (1980) The model-theoretic argument against realism. *Philosophy of Science* 47: 69–81.
Papineau, David (1996) Theory-dependent terms. *Philosophy of Science* 63: 1–20.
Psillos, Stathis (1999) *Scientific Realism: How Science Tracks Truth*. London: Routledge.
(2000) Carnap's 'Theoretical Concepts in Science'. *Studies in History and Philosophy of Science,* pt. A, 3: 151–172.
(2006) Ramsey's Ramsey-sentences. In M. C. Galavotti (ed.) *Cambridge and Vienna: Frank P. Ramsey and the Vienna Circle*. Dordrecht: Springer, pp. 67–90.
(2008) Carnap and incommensurability. *Philosophical Inquiry* 30: 135–156.
Psillos, Stathis and Christopoulou, Demetra (2009) The a priori: Between conventions and implicit definitions. In N. Kompa, C. Nimtz, and C. Suhm (eds.) *The A Priori and Its Role in Philosophy*, Paderborn: Mentis, pp. 205–220.
Putnam, Hilary (1962) What theories are not. In E. Nagel, P. Suppes, and A. Tarski (eds.) *Logic, Methodology, and Philosophy of Science*, Palo Alto, CA: Stanford University Press, pp. 240–251.

(1983) *Realism and Reason*, vol. III of *Philosophical Papers*. Cambridge University Press.
Raatikainen, Panu (2007) Theories of reference and the philosophy of science. Paper presented at EPSA07: 1st Conference of the European Philosophy of Science Association Madrid, 15–17 November.
Sankey, Howard (2009) Scientific realism and the semantic incommensurability thesis. *Studies in History and Philosophy of Science* 40: 196–202.
Schlick, Moritz ([1932] 1979) Form and content: An introduction to philosophical thinking. In Moritz Schlick, *Philosophical Papers*, vol. II, Dordrecht: D. Reidel, pp.285–369.
Turner, Edward (1835) *Elements of Chemistry*, Philadelphia: Desilver Thomas & Co.
Winnie, John (1967) The implicit definition of theoretical terms. In *British Journal for the Philosophy of Science* 18: 223–229.
(1970) Theoretical analyticity. In R. Cohen and M. Wartofsky (eds.) *Boston Studies in the Philosophy of Science*, vol. VIII. Dordrecht: Reidel, pp. 289–305.
Ure, Andrew ([1820] 1828) *Dictionary of Chemistry*, 3rd edn. London: Thomas Tegg.