E-LOGOS – Electronic Journal for Philosophy 2020, Vol. 27(2) 57–70
ISSN 1211-0442 (DOI 10.18267/j.e-logos.479), Peer-reviewed article
Journal homepage: e-logos.vse.cz

In Defense of Semantic Externalism

Panu Raatikainen[1]

Abstract: The most popular and influential strategies used against semantic externalism and the causal theory of reference are critically examined. It is argued that upon closer scrutiny, none of them emerges as truly convincing.

Keywords: externalism, meaning, reference, natural kinds.

[1] *Degree Programme in Philosophy, Pinni B4147, FIN-33014 Tampere University, Finland, e-mail: panu.raatikainen@tuni.fi*

1 Introduction

Semantic externalism maintains, against the more traditional internalist[2] views on meaning, that the meaning[3] of a referring expression may *not* always and exhaustively be determined by a language user's mental states – by what is, so to say, in the language user's mind – but that the social and physical environment may also play some role in it; in other words, that *some* words mean, or refer, in virtue of relations that are *partly* external to the mind.[4] Words which refer to *natural kinds* are commonly viewed as standard examples of such expressions.

This view has won many adherents, but it has also been vigorously attacked by numerous able philosophers. Many seem to be confident that it has been rebutted for good. It is obviously not possible discuss all of the dozens of critical reactions in the literature here. However, there is a certain convergence in these responses; the same reply strategies tend to recur over and over again in the critical literature.[5] In this paper, I shall critically examine what are apparently the most influential critique strategies aiming to undermine externalist arguments. There have been some scattered externalist responses to them here and there in the literature, but my aim in this paper is to bring these critique strategies together and provide a more systematic and comprehensive critical discussion of them. At the same time, the aim is to clarify what more exactly semantic externalism is, and what it is not.

2 Classical Arguments for Semantic Externalism

The debate has focused especially on Putnam's classic arguments in favor of externalism, in his seminal paper "Meaning of 'Meaning'" (1975). Although well known, let us review them briefly for later reference.[6] To begin with, there are two different factors in externalism, which Putnam has called the division of linguistic labor and the contribution of the environment;

---

[2] As far as I know, the label "internalism", in this context, originated with Searle (1983). Accordingly, it has been natural to call the new opposite view "externalism", which has become a common label.

[3] That is, meaning understood as something that determines reference. I shall not consider here the possibility of giving up this assumption. (Farkas 2006, for example, suggests such a response.) Such a line of rejoinder would deserve a discussion of its own. Let it suffice here to note that this would certainly be a strong deviance from the traditional view on meaning. I think there is a threat here that the whole disagreement becomes largely verbal.

[4] Accordingly, the strength of the externalist thesis should not be exaggerated; as Devitt, a leading externalist writes: "[I]t is not a consequence [of Putnam's slogan] that no aspect of meaning is in the head. The point of the slogan is simply to deny that meanings are entirely in the head. In my view, the meaning of a term is likely to involve many psychological states ... the slogan emphasizes that extra-cranial links to reality are also necessary to meaning." (Devitt 1990, p. 83).

[5] Thus one reads, for example, in an encyclopedia entry by Bach (1998) that "[t]hese thought experiments [of Twin Earth etc.] have met with considerable enthusiasm but also with neglected criticism"; one is then referred to (Unger 1983), (Bach 1987), and (Crane 1991). Also, e.g., Segal (2000) refers largely to these same sources and recycles much the same arguments (see also Segal 1999). (Zemach 1976), (Mellor 1977), (Crane 1991) and excerpts from (Searle 1983) are reprinted in The Twin Earth Chronicles (Pessin & Goldberg 1996). Further, both Kallestrup (2012), in the section on internalist rejoinders (3.2), and Smith (2013), in the section "Initial problems with the Twin Earth thought experiments" (3), refer largely to these same sources (Kallestrup also refers to causal descriptivism; see Section 11 below). Thus I think that there is a rather wide agreement on what the most important philosophical arguments against externalism are.

[6] Certainly also others, e.g., Donnellan, Kripke, Burge and Devitt, have presented important arguments in support of externalism. The critical reactions I shall consider below have, however, focused mainly on Putnam's arguments. For this reason, I shall focus in particular on Putnam's discussion.

sometimes the corresponding variants of externalism are called social externalism and physical externalism, respectively. The arguments for these variants are somewhat different.

Putnam illustrated social externalism e.g. with the following example: Putnam admits that he himself cannot tell an elm from a beech tree. Even so, he maintains that the extension of "elm" in his idiolect is the same as the extension of "elm" in anyone else's idiolect, and similarly with "beech". Not everyone in the language community has to be able to recognize elms, or know the necessary and sufficient conditions for being an elm – or a tiger, water, aluminium, or gold, etc. (i.e., so-called natural kinds). Most people acquire such a word without acquiring a method of recognizing reliably whether something really is in the extension of the word, or knowing the necessary and sufficient conditions for belonging to the extension of the word. Nevertheless, they can use the word and refer with it successfully. Putnam suggested that the average language user can defer to experts. This is the basic idea of social externalism, or the division of linguistic labor.[7]

Furthermore, it has also been argued that successful referring is possible even in a situation where *nobody* in the linguistic community knows the necessary and sufficient conditions for belonging to the extension of an expression. Certainly the most famous argument for this kind of externalism is Putnam's Twin Earth thought experiment. Putnam invites us to imagine that somewhere, far, far away, there is a planet very much like Earth; let us call it Twin Earth. We may even assume that every one of us has a Doppelgänger there and that languages similar to ours are spoken there. There is, however, a peculiar difference between Earth and Twin Earth: the liquid called "water" in Twin Earth is not H_2O but a totally different liquid whose chemical formula is very long and complicated; we may abbreviate it as XYZ. It is assumed that it is indistinguishable from water in normal circumstances; it tastes like water and quenches thirst like water, lakes and rivers of Twin Earth contain XYZ, it rains XYZ there, etc. Putnam next assumes that we roll the time back to, say, 1750, when chemistry had not been developed on either Earth or Twin Earth. At that time nobody would have been able to differentiate between XYZ and H_2O. In this case it is not possible to depend on experts. Nobody is in the know. But still, Putnam contends, the extension of "water" was just as much H_2O on Earth, and the extension of "water" was just as much XYZ on Twin Earth.

Consider thus the use of "water" by Oscar on Earth and by Twin Oscar on Twin Earth in 1750. Oscar and Twin Oscar are exact duplicates, so they are stipulated to have the same internal mental states. Nevertheless, they refer, according to Putnam, to different substances by their tokens "water". Hence, their mental states do not determine the meanings of their words. Putnam expressed this conclusion by his famous slogan: "Cut the pie any way you like, 'meanings' just ain't in the head!" This is what is meant by the claim of physical externalism that meaning is partly determined by the environment.

To be more precise, an important part of Putnam's more exact argument is the qualification that the mental states in question are assumed to be "narrow" (Putnam 1975, p. 139; cf. Devitt 1990). For it is all too easy to find "wide" mental states which *can* determine the extensions of expressions: Consider, say, Oscar's intention to refer, by his expression "water", to water.

---

[7] *I am not myself unconditionally happy with this part of Putnam's view. I believe that reference depends on earlier users of the word in the relevant causal-historical chain of reference-borrowing (along the lines that Kripke suggested) – and not on the present experts. (This is also argued by Kripke in a less known little paper (Kripke 1986).) Obviously, however, an introducer of a word may sometimes be an expert.*

Consequently, the interesting claim worth serious consideration (and not trivially false) is that the language user's *narrow* mental states do not always suffice to determine reference.

However, providing an altogether satisfactory definition of "narrow state" has turned out to be difficult. Putnam's original, intuitive definition, according to which a state is narrow if it does not presuppose the existence of any individual other than the subject to whom that state is ascribed, has its problems (see, e.g., Burge 1982, Williamson 2006). The following revised definition, offered by Williamson (2006), apparently avoids them:

Call a state *S narrow* if and only if whether an agent is in *S* at a time *t* depends only on the total internal qualitative state of *S* at *t*, so that if one agent in one possible situation is internally an exact duplicate of another agent in another possible situation, then the first agent is in *S* in the first situation if and only if the second agent is in *S* in the second situation.

But be that as it may, I shall not pay that much attention to the exact details of the notion of narrowness in what follows, as the responses to externalism I am focusing on below do not question the possibility of giving a satisfactory characterization of "narrow". Moreover, like Williamson I am inclined to think that were it to turn out that it is not at all possible to draw a non-arbitrary line between the internal and the external, that would be more a problem for traditional internalism rather than for externalism. This is because it is internalism for which the internal–external distinction has some theoretical importance. Hence, we may assume, for the sake of an argument, that a principled boundary can be drawn. (cf. Williamson 2006.)

3 The Causal Theory of Reference

A common part of semantic externalism is an alternative positive view on reference[8] – more a sketch or a "picture" than a developed theory – known as "the causal theory of reference" (Donnellan, for example, prefers the label "the historical theory of reference"; it is also sometimes called "the causal-historical theory of reference", or simply "the new theory of reference"). This account can be especially attributed to Kripke (1980).[9] Putnam and others have largely accepted it (see, e.g., Putnam 1975, Donnellan 1974, Devitt 1981).

Roughly, the idea is that an expression refers to whatever is appropriately causally linked to the language user. The picture has two parts: a theory of initial fixing of reference, and a theory of reference borrowing. First, it is suggested that a referring expression is *typically* introduced, in a "baptism" or a dubbing event, in the perceptual contact with the referent or a sample of the kind (although an expression can also be introduced with the help of mere description, without any perceptual contact). Second, other language users not present at the name-giving occasion acquire the word from those present at the dubbing, still others from the former, and so on. Later language users gain the ability to refer with the expression in virtue of an appropriate causal-historical chain going back to the introduction of the expression. Language users may be largely ignorant of this chain. Nevertheless, they can successfully refer with the expression. This is the idea of reference borrowing.

The causal-historical picture harmonizes well with the general externalist view, and helps to explain in part how meanings can be external to the mind: the causal-historical chain which

[8] This is not necessarily so; the negative arguments by Putnam and others against internalism do not essentially assume any detailed positive theory of reference.

[9] Kripke himself developed this "theory" as a part of his attack on the description theory of reference (unlike Putnam, he was not concerned with internalism as such).

determines reference of an expression typically extends back in history and goes beyond what a particular language user knows or what is in her mind.

These are the main lines of semantic externalism and the causal theory of reference. Let us now look at what appear to be the most influential attempts to rebut them.

4 The Common Concept Strategy

One very popular countermove against externalism has been what is sometimes called "the common concept strategy". That is, a repeated reply to Putnam's Twin Earth thought experiment is the suggestion that why not to simply say that our word "water" had (in the scenario in which Twin Earth exists) – at least when nobody knew the chemical constitution of water – in its extension both H_2O and qualitatively similar XYZ.[10]

Now, one may flirt with this reply strategy in a particular case (e.g., with "water" and XYZ), but all that externalism really needs is the possibility of a case where everything appears the same but nevertheless two speakers talk about different things. In the thought experiment, XYZ can be *stipulated* to be a substance *different* from water and only superficially similar to it. And denying the very possibility of such a case apparently amounts to the claim that there can be no difference in substances without some observable difference – that everything that looks like the same substance really is the same substance. But the latter is a strong and controversial empiricist thesis which is, at best, in need of a substantial defense and certainly can't be taken for granted (see McCulloch 1992, 2003).[11] I think that even more is true: I suspect that the large majority of philosophers today agree that such a radical empiricism is as dead as a philosophical view can be.

Note also that this strategy seems to implicitly make the quite implausible and unprincipled assumption that in the scenario in which Twin Earth exists, when we finally learned the chemical constitution of water, the extension and consequently the meaning of the word "water" changed radically (it used to have both H_2O and XYZ in its extension but afterwards only H_2O).[12] But wouldn't it be much more plausible to say that we just learned what the constitution of water is – and assume that the meaning (or at least the extension) of "water" remained stable?[13] For if one begins to assume that there is a change in the meaning of a term whenever the theory associated with the term changes, one is dangerously sliding towards radical meaning variance, incommensurability and conceptual relativism á la Kuhn and Feyerabend. And anyone with even a modest faith in the rationality of science should not start going down that road.

Or, to change the example, let us consider gold (the following example is taken from Putnam 1994). What *chrysos* (gold) was in ancient Greece was not simply determined by the properties ancient Greeks believed gold to have. For otherwise it would have made no sense for an ancient Greek to ask himself, "Is there perhaps a way to tell that something isn't really gold, even when it appears by all standard tests to be gold?" But this is precisely the question Archimedes put to

---

[10] See e.g., Zemach 1976; Mellor 1977; Searle 1983, p. 203; Bach 1987, p. 276; Crane 1991; Segal 2000, p. 19; cf. Farkas 2006; Wikforss 2008; Haukioja 2017.

[11] Interestingly, Crane has accepted this reply and has given up the common concept strategy (see Crane 2001, p. 123).

[12] Apparently the meaning would have changed also in the actual scenario, even without any Twin Earth.

[13] Indeed, it is unclear what chemical constitution we then learned, if the common concept strategy was true – not that of water, as it was then understood, for "water" meant both H_2O and XYZ.

himself, with a celebrated result. Archimedes's inquiry would have made no sense if he had not had the idea that something might appear to be gold (might pass the current tests for "chrysos"), while not actually having the same nature as the paradigm examples of gold (see Putnam 1994, p. 443–444). But apparently the common concept strategy entails that a case such as this is impossible.

Finally, consider the following modified Twin Earth story told by Kim Sterelny (1983). This variant makes the nowadays widely accepted assumption that it is possible for two individuals to be in a biochemically different but psychologically identical state (so-called multiple realizability). In this version, H_2O, if any were to exist on Twin Earth, would be toxic and foul tasting to Twin Earthians, and conversely for XYZ and us. In this scenario, presumably nobody would suggest that both H_2O and XYZ belong to the extension of our word "water" (and also of "water" in Twin Earth), as the common concept strategy suggests. But if Putnam ever could plausibly claim that Oscar and Twin Oscar were in the same narrow psychological state in respect to "water" he still can. This premise of Putnam's argument has not been challenged, and nothing essential to the Twin Earth argument has been changed. Accordingly, psychological states, internally defined, do not determine reference (see Sterelny 1983).

In sum, it should be abundantly clear by now that the common concept strategy cannot rebut the Twin Earth argument in favor of externalism. It also leads to various implausible consequences.

5 The Incomplete Understanding Strategy

Recall Putnam's example him being unable to tell an elm from a beech tree. As a reply to this, it has been frequently suggested[14] that in such cases, one does *not* have *a full* or *complete understanding* of the word (say, "elm") – that one does not know the meaning, or has not grasped the meaning, completely. This rejoinder, however, is both unnatural and irrelevant.

The proposed response is unnatural because it entails that the average language users do not understand their own native language – not even its everyday basic part which consists of such familiar words as "elm", "tiger", "gold", etc. Only experts would understand words such as these. More importantly, this "incomplete understanding strategy" (as one might call it) is not actually relevant, because the real issue is whether people are able to successfully refer with an expression or not, even if they perhaps only "partially grasp" the meaning of the expression. It is indeed a key claim of semantic externalism that often language users do *not* literally know the meanings of their language, if it is agreed that meaning determines reference – and hence also that the knowledge of meaning entails the knowledge of reference (cf. Putnam, 1988, p. 32) – i.e., that there is a sense in which we do *not*, for the most part, know what we mean (see Devitt 1981, p. 20; cf. Raatikainen 2010).[15]

Consequently, externalism and its opponents need not disagree here. Externalism, however, adds the further claim that people may nevertheless be able to refer successfully (e.g. to elms,

---

[14] See, e.g., Searle 1983, p. 201; Bach 1987, p. 287–290; Crane 1991, Stanley 1999.

[15] This does not amount to the absurd claim that we do not understand our native language. Rather, the point is to distinguish the ordinary understanding of one's native language, or normal linguistic competence, as the ability to use it in discourse and successfully refer with its expressions, from the unrealistically demanding idea of the literal knowledge of the meanings of one's language (where meaning is something that determines reference).

and only elms) even in such situations. The opponents of externalism need to deny this. Such a denial, however, seems to take one back to something like the common concept strategy, e.g., to the claim that "elm" in Putnam's idiolect has in its extension both elms and beeches. But we have already found this strategy wanting.[16] Alternatively, it might be contended that a word does not refer, in such circumstances, to anything at all – but that seems excessively radical, especially given how common such situations of ignorance are.

It just appears to be a fact of life that people do manage to use words to refer even though they are quite ignorant, for example, when they request more information about the referent. For instance, if an ignorant like Putnam asks, "What do the leaves of elms more specifically look like?", it seems plausible that by "elm" in this question he is quite successfully referring to elms, and only elms.

6 Refuting Counterexamples?

Some critics, e.g., Unger (1983) and Searle (1983, p. 239), seem to suggest that refuting counterexamples exist for the causal theory of reference. Such a criticism, though, assumes that the causal theory of reference is a general theory of how expressions refer. However, it was never claimed – by Kripke, Putnam, or Devitt, for example – that *all* expressions, or even all proper names, refer along the lines of the causal theory of reference.

That *some* expressions really are, in a sense, descriptive was admitted from the beginning: e.g., Kripke gave as an example "Jack the Ripper",[17] Putnam "vixen", and Devitt "pediatrician". More exactly, descriptions can sometimes, even according to the causal theory of reference, play an essential role in the *introduction* of an expression. (Even in such a case, reference borrowing does not require that description to be transmitted with the expression.) Devitt (1981) even systematized this and generalized Donnellan's distinction between referential and attributive uses of descriptions to apply to all sorts of expressions, including names. In this terminology, typical names are referential, but "Jack the Ripper," for example, is attributive.

Therefore, some alleged counterexamples cannot refute externalism, which only maintains that the meaning of an expression may *sometimes* go beyond what is in a "speaker's mind", or that *not all* referring expressions are synonymous with some associated descriptions (or clusters of descriptions).

7 The No Failures of Reference Objection

It has been repeatedly claimed that semantic externalism, or the causal theory of reference, is unable to account for reference failures.[18] That is, it is proposed that the causal theory of reference cannot explain the cases where one has ended up with the conclusion that certain postulated entities did not exist after all. The once postulated planet Vulcan and phlogiston, a hypothetical substance of fire, are popular examples. The reasoning commonly proceeds along the following lines: Certainly those who first introduced the non-referring term were causally

---

[16] Thus, what happens if there are no experts, no one to whom defer the reference – like with "water" in 1750? This strategy entails that either nobody understands the expression or that there must later occur a radical meaning change. Neither alternative is plausible.

[17] Kripke adds, "But in many or most cases, I think the thesis [descriptivism] is false" (Kripke 1980, p. 80).

[18] See, e.g., Enc 1976, Nola 1980, Kroon 1985, Psillos 2000, p. 290; Niiniluoto 1999, p. 126; Bird 2000, p. 185; this complaint seems to be especially popular among the philosophers of science.

interacting with something that caused the phenomenon which the entities were postulated to explain. For example, it is now universally agreed that "phlogiston" failed to refer to anything real. But there was something, namely oxygen, present in combustion. Therefore, the argument continues, if the causal theory of reference were correct, "phlogiston" should refer to oxygen rather than fail to refer. And that is implausible.

But this is too hasty. First, the advocates of semantic externalism have, after the earliest pioneering papers in the tradition, discussed the issue of reference failure in some detail.[19] And second, once it is admitted that that *some* expressions are in a sense descriptive, or "attributive" (see above), and that many clear-cut examples of non-referring expressions are arguably such, nothing prevents an externalist from accounting *their* possible failure to refer along traditional descriptivist lines. It should be in fact quite clear that names such as "Vulcan" or "phlogiston", for example, were *not* introduced in perceptual confrontation with their referent. Consequently, they must have been introduced with a help of descriptions (Devitt has called such naming ceremonies "abnormal"). The causal-perceptual account of baptism only applies to observational natural kinds and entities for which there is a perceptual contact to the referent in the initial name-giving. Critics of the causal theory of reference repeatedly ignore this.

Consider, thus, for instance, Kripke's own example of "Jack the Ripper," which was perhaps introduced, even according to Kripke, with a description along the lines "the man, whoever committed all these murders, or most of them" (Kripke 1980, p. 79). If it, however, turned out that each one of those murders was committed by a different person or that there were no murders at all but that these were all improbable accidents, one would likely conclude that "Jack the Ripper" does not refer to anything real.

Although a typical name introduction event does involve perception and a causal interaction with the bearer of the name, the *causal chain* is, in the Kripkean causal-historical account, primarily a chain of communication between the earlier and later *uses* and *users* of the name, and mostly concerns reference borrowing. It does not require the bearer of the name to be a causal *relatum*. Hence the criticism at issue is based on a misinterpretation of the theory. In sum, reference failures present no serious problem for the causal-historical picture of reference (for natural kinds terms and reference failure, see also Section 9 below).

8 Thought Experiments with Alternative Conclusions

Unger (1983) has devised some ingenious variations on Putnam's Twin Earth thought experiment that seem to support contrary intuition.[20] Many of them, and arguably the most puzzling ones, are based on some radical but unnoticed change in the environment. For example, imagine that all the water (i.e., H_2O) on Earth was replaced overnight (say, secretly by aliens) by XYZ; what would the extension of "water" here on Earth be after, say, 100 years?[21] The positive account of Kripke, Putnam and others (that is, the causal theory of reference in its original form) seems to require that the extension would be only H_2O. But intuitively, this does not seem at all clear: it looks as if "water" would sooner or later switch its reference to XYZ (such scenarios are sometimes called "slow switching" cases in the literature).

---

[19] See, e.g., Donnellan 1974, Devitt 1981, Braun 1993, Salmon 1998, Reimer 2001.

[20] Also Bach, for example, refers to them; see Bach 1987, 276–277; cf. Bach 1998.

[21] This is my own example, not Unger's, but I think that it captures fairly the basic idea of many of his cases.

Now the critical literature on externalism has a regrettable tendency to focus solely on the earliest statements of semantic externalism and the causal theory of reference, and totally ignore its later developments. Perhaps the most important further refinement of the causal-historical theory of reference is Devitt's idea of "multiple grounding". That is, the early, sketchy formulations of the causal theory of reference seem indeed to imply that the reference of an expression can never change. But that is implausible, as Evans (1973) clearly pointed out. Devitt, however, has suggested that it is not only the initial dubbing or baptism which determines the reference, but that a word typically becomes multiply grounded in its bearer in other uses of the word relevantly similar to a dubbing; that is, they involve the application of the word to the object in direct perceptual confrontation with it (see Devitt 1981, 57-58; Devitt & Sterelny 1999, 75-76).[22] This idea makes reference change possible and enables one to reply, not only to Evans' initial reference change worry, but also to most of Unger's alleged counterexamples to the causal theory of reference (Unger 1983). Critics of externalism tend to ignore this important improvement.

9 Wrong Generalizations: the *Qua* Problem

Semantic externalism and the causal theory of reference as applied to kind terms has been often criticized as follows: A sample will usually be a member of many kinds. For example, a particular tiger is simultaneously, say, an Indochinese tiger, a tiger, a feline, a mammal, and an animal, as well as a predator and a striped animal. So how can a *general term* such as "tiger" be introduced? If it happens through an initial baptism in the contact with a sample, as the causal theory of reference seems to suggest, how can one rule out incorrect kinds of generalizations?[23] This is the so-called *qua* problem (see Devitt & Sterelny 1999, 72–75, 90–93).

In fact, however, it has long been recognized among advocates of the causal theory of reference that, especially in the case of general terms, the introduction of a word must involve *some* descriptive content (see, e.g., Sterelny 1983, Devitt & Sterelny 1987 and 1999, and Stanford & Kitcher 2000). Stanford and Kitcher (2000) in particular have substantially improved Putnam's original account of the reference of natural kind terms. Roughly, in their approach, there is a whole range of samples (not only a single sample), also a range of foils, and some associated properties involved in the introduction of a natural kind term. This shows how one can rule out the wrong kind of generalization (at least many of them), and it also shows how an apparent natural kind term can fail to refer to anything.

According to the approach of Stanford and Kitcher, term introducers make stabs in the dark: they see some observable properties that are regularly associated, and *conjecture* that *some* underlying property (or "inner structure")[24] figures as a common constituent of the total causes of each of the properties. This conjecture may be incorrect, in which case the term may fail to refer. But if it is correct, one can exclude incorrect generalizations and fix the reference in the

---

[22] In fact, Devitt first proposed this refinement already in (Devitt 1974). Putnam (2001) comments on it approvingly. Also, Kripke (1980, p. 163) acknowledges the need for such a refinement, but seems to be unaware of Devitt's suggestion.

[23] See, e.g., Papineau 1979, Dupre 1981, Crane 1991, Segal 2000.

[24] Or history, which may likewise be opaque for the introducer of the word: in the case of biological species, for example, the evolutionary lineage.

intended way to the set of things that share that underlying property, belong to the same species, etc.

In such a situation, it may perhaps sometimes remain indeterminate whether or not some borderline cases belong to the extension of a term (e.g., heavy water, often mentioned by the critics of semantic externalism), and there may be some room for conventional choice, but this is not relevant to the fundamental issue.[25] Multiple grounding (see above) may also play a role in such cases. Be that as it may, only superficially similar but internally *radically* different objects or substances (as XYZ in Putnam's thought experiment, for example, is *stipulated* to be) simply do not belong to the extension, and this is sufficient for the general argument in favor of semantic externalism.

10 Causal Descriptivism and Metalinguistic Descriptivism

The causal theory of reference was originally developed by Kripke as an alternative to the description theory of reference, or descriptivism, in short. According to the latter theory, the reference of an expression is determined by a description (or a cluster of descriptions) that the language user analytically associates with the expression. Kripke (1980), Donnellan (1970) and others have argued against such views by emphasizing that language users are often much too ignorant and erring to have the kind of identifying knowledge required by descriptivism (see e.g. Devitt & Sterelny 1999). Putnam's Twin Earth argument can be interpreted as a particularly powerful variant of an argument from ignorance.

Now, however, a new form of the description theory of reference hopes to restore the old order. Its ingenious idea is to mimic the new causal-historical theory of reference in its descriptions; hence its name "causal descriptivism" (CD). Such a form of descriptivism has been suggested by David Lewis (1984), Frederick Kroon (1987), and Frank Jackson (1998), for example, and seems to be enjoying some popularity.

One may express the basic idea of causal descriptivism schematically as follows: speakers associate a name "*N*" with a description of the form[26]

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

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

Another somewhat related recent variant of descriptivism, favored for example by Searle (1983), Bach (1987), and Katz (1990, 1994), is called "nominal descriptivism" or "meta-linguistic descriptivism" (henceforth MD). (MD is sometimes taken as a form of causal descriptivism; e.g., by Kroon). It is also built into the two-dimensional semantic framework of Chalmers (2002). According to MD, a description which suits the purpose for a proper name "*N*", is simply of the form

---

[25] Hendry has developed much more detailed account, focusing especially on the development of chemistry, of how more exactly such situations should be analyzed (see e.g. Hendry 2010). He argues, in my mind convincingly, that even in the case of heavy water, there are principled reasons for counting it as water, and that the choice is not arbitrary.

[26] Some variants: "The thing which is a bearer of 'N'" (Katz); "The bearer of 'N'" (Bach); "The object called 'N' in my linguistic community or at least by those from whom I got the name" (Searle); such small differences do not affect what I'll say below.

The thing to which "*N*" refers.

Now philosophers proposing such new variants of descriptivism may have had various differing philosophical aims in mind when putting them forward. This is not the place to evaluate whether they have succeeded in achieving some of their goals with the help of such causal or metalinguistic descriptions.[27] The crucial question, in the present context, is whether CD or MD can be used to salvage *internalism* – as one often seems to think they can.

To begin with, one should note that these strategies are actually incoherent with some of the other strategies discussed above, for example, with the incomplete understanding strategy and perhaps also with the common concept strategy. That is, whereas the latter put a considerable epistemic burden on the language user in order for her to be able to successfully refer, causal descriptivism or metalinguistic descriptivism make successful referring stunningly easy, even trivial. Some, such as Searle and Bach, appeal to all these strategies, but it would be certainly better not to do so.

Further, it seems that CD and MD are most often proposed only as theories of proper names. Semantic externalism, in contrast, is at least as much a theory of the meaning and the reference of general terms (and natural kind terms in particular). Hence, CD and MD, so understood, are from the start quite powerless to undermine externalism. If, on the other hand, CD and MD are also intended as theories of general terms, one must face the problem of whether they are meant to be theories of meaning. For externalism is obviously a thesis about meaning. But it is fair to say that, with respect to general terms in particular, CD and MD are quite implausible as theories of meaning.[28]

Let us, however, only assume that CD and MD are used merely as a theory of what determines reference, not as a theory of meaning. We may thus ask whether they can be at least used to secure the weaker internalist doctrine that reference is determined by a language user's narrow mental states. It seems as if they cannot do even that. (For simplicity, I shall consider only proper names – but the moral clearly generalizes to general terms.)

Imagine that both Oscar and Twin Oscar associate with their respective tokens of "Aristotle" the property of being the referent of "Aristotle," as MD suggests. Now presumably such a thought of Oscar is about our Earthly Aristotle (Aristotle is the referent of "Aristotle" in his linguistic community), whereas Twin Oscar's thought is about Twin Aristotle who once lived in Twin Earth (Twin Aristotle is the referent of "Aristotle" in his linguistic community).[29] Hence, apparently the resulting mental state (state resulting from the association of the description with the name) is not the same for Oscar and Twin Oscar, in other words, is not narrow, as was required.[30] An analogous argument applies to CD. Consequently, CD and MD are entirely powerless as weapons against semantic externalism. They cannot be used to save internalism.

---

[27] (Raatikainen 2020) includes a rather comprehensive critical discussion of such views; see also (Everett 2005).

[28] For more about such problems, see Raatikainen 2006, 2020; Everett 2005.

[29] I take it that the alternative that "Aristotle" refers, in both Oscar's and Twin Oscar's idiolect, somehow disjunctively to both Aristotle and Twin Aristotle is blatantly absurd.

[30] Recall our improved definition (due to Williamson) of "narrow state" from the end of Section 2.

11 Conclusion

We have examined above the most popular and influential strategies used against semantic externalism. Upon closer scrutiny, none of them emerges as truly convincing. None of them are successful in undermining externalism. The arguments in favor of semantic externalism, on the other hand, have considerable force. At least, if someone wants to evade the externalist conclusion, the burden is on the opponent of externalism to provide other, more convincing arguments.

Acknowledgements

Different versions of this paper have been presented over time in Barcelona, Stockholm, Vietri, and Turku. I am grateful for the audiences on all these occasions for many useful comments. I am also indebted to discussions, comments, and correspondence on the themes of the paper with Tim Crane, Michael Devitt, and Tim Williamson. Naturally, I am solely responsible for the content of the paper.

There is certain overlap between this paper with my papers (Raatikainen 2020) and (Raatikainen 2021), but all three have quite different foci and in my view, they usefully complement each other. A version of this paper has appeared in Finnish as (Raatikainen 2019).

References

Bach, Kent (1987). *Thought and Reference*. Oxford: Oxford University Press.

Bach, Kent (1998). Content: wide and narrow. In E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge.

Bird, Alexander (2000). *Thomas Kuhn*. Chesman: Acumen.

Braun, David (1993). Empty names. *Noûs* 27, 449–469.

Burge, Tyler (1979). Individualism and the mental. *Midwest Studies in Philosophy* 4, 73–121.

Burge, Tyler (1982). Other bodies. In A. Woodfield (ed.), *Thought and Object: Essays on Intentionality*. Oxford: Oxford University Press, 97–120.

Crane, Tim (1991). All the difference in the world. *The Philosophical Quarterly* 41, 1–25.

Crane, Tim (2001). *Elements of Mind*. Oxford: Oxford University Press.

Devitt, Michael (1974). Singular terms. *Journal of Philosophy* 71, 183–205.

Devitt, Michael (1981). *Designation*. New York: Columbia University Press.

Devitt, Michael (1990). Meanings just ain't in the head. In G. Boolos (ed.), *Meaning and Method: Essays in Honor of Hilary Putnam*. Cambridge University Press, 79–104.

Devitt, Michael and Kim Sterelny (1987). *Language and Reality*. Oxford: Basil Blackwell.

Devitt, Michael and Kim Sterelny (1999). *Language and Reality*, Second Edition. Oxford: Blackwell.

Donnellan, Keith (1970). Proper names and identifying descriptions. *Synthese* 21, 335–358. Reprinted in D. Davidson & G. Harman (eds.), *Semantics of Natural Language*. Dordrecht: Reidel, 1972, 356–79.

Donnellan, Keith (1974). Speaking of nothing. *Philosophical Review* 83, 3–31.

Dupré, John (1981). Natural kinds and biological taxa. *Philosophical Review* 90, 66–90.

Enç, Berent (1976). Reference of theoretical terms. *Noûs* 10, 261–282.

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

Everett, Anthony (2005). Recent defenses of descriptivism. *Mind & Language* 20, 103–139.

Farkas, Katalin (2006). Semantic internalism and externalism. In E. Lepore and B.C. Smith (eds.), *The Oxford Handbook of Philosophy of Language*, Oxford: Oxford University Press, 323–40.

Haukioja, Jussi (2017). Internalism and externalism. In B. Hale, C. Wright & A. Miller (eds.), *A Companion to the Philosophy of Language*, 2nd edition, Vol. II. Oxford: Wiley Blackwell, 865–880.

Hendry, Robin (2010). The elements and conceptual change. In Helen Beebee & Nigel Sabbarton-Leary (eds.), *The Semantics and Metaphysics of Natural Kinds*. New York: Routledge, 137–158.

Jackson, Frank (1998). Reference and description revisited. *Philosophical Perspectives* 12, 201–218.

Kallestrup, Jesper (2012). *Semantic Externalism*. London and New York: Routledge.

Kripke, Saul (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Kripke, Saul (1986). A problem in the theory of reference: the linguistic division of labor and the social character of naming. *Philosophy and Culture* (Proceedings of the XVIIth World Congress of Philosophy), Montreal: Editions Montmorency, 241–247.

Kroon, Frederick (1985). Theoretical terms and the causal view of reference. *Australasian Journal of Philosophy* 63, 143–166.

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

Lewis, David (1984). Putnam's paradox. *Australasian Journal of Philosophy* 62, 221–36.

McCulloch, Gregory (1992). The spirit of twin earth. *Analysis* 52, 168–174.

McCulloch, Gregory (2003). *The Life of the Mind*. London: Routledge.

Mellor, D. H. (1977). Natural kinds. *British Journal for the Philosophy of Science* 28, 299–312.

Nola, Robert (1980). Fixing the reference of theoretical terms. *Philosophy of Science* 47, 505–531.

Papineau, David (1979). *Theory and Meaning*. Oxford: Clarendon Press.

Pessin, Andrew and Sanford Goldberg, (eds.) (1996). *The Twin Earth Chronicles. Twenty Years of Reflection on Hilary Putnam's "The Meaning of 'Meaning'"*. New York: M.E. Sharpe.

Putnam, Hilary (1975), The meaning of "meaning". In Keith Gunderson (ed.), *Language, Mind and Knowledge. Minnesota Studies in the Philosophy of Science* VII. Minneapolis: University of Minnesota Press, 1975, 131–193.

Putnam, Hilary (1994). Why functionalism didn't work. In H. Putnam: *Words and Life*. Cambridge: Harvard University Press, 441–459.

Putnam, Hilary (2001). Reply to Michael Devitt. *Revue Internationale de Philosophie* 4/2001, No 218, 495–502.

Raatikainen, Panu (2006). Against causal descriptivism. *Mind & Society* Vol 5, No 1, 78–84.

Raatikainen, Panu (2010). The semantic realism/anti-realism dispute and knowledge of meanings. *The Baltic International Yearbook of Cognition, Logic and* Communication Vol. 5, October 2010, 1–13.

Raatikainen, Panu (2019). Semanttisen eksternalismin puolustus. *Ajatus* 76, 11–36.

Raatikainen, Panu (2020). Theories of reference: what was the question? In A. Bianchi (ed.), *Language and Reality from a Naturalistic Perspective: Themes from Michael Devitt*. Cham: Springer, 69–103.

Raatikainen (2021). Natural kind terms again. *European Journal for Philosophy of Science* (forthcoming).

Reimer, Marga (2001). The problem of empty terms. *Australasian Journal of Philosophy* 79, 491–506.

Salmon, Nathan (1998). Nonexistence. *Nous* 32, 277–319.

Searle, John (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

Segal, Gabriel (2000). *A Slim Book About Narrow Content*, Cambridge MA: MIT Press.

Segal, Gabriel (1999). Twin Earth. In R.A. Wilson & F.C. Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, MA: MIT Press, 850–52.

Smith, Basil (2013). Internalism and externalism in the philosophy of mind and language. *Internet Encyclopedia of Philosophy*. http://www.iep.utm.edu/int-ex-ml/

Stanley, Jason (1999). Understanding, context-relativity, and the description theory. *Analysis* 59: 14–18.

Sterelny, Kim (1983). Natural kind terms. *Pacific Philosophical Quarterly* 64, 110–25.

Unger, Peter (1983). The causal theory of reference. *Philosophical Studies* 43, 1–45.

Zemach, Eddy (1976). Putnam's theory of reference of substance terms. *Journal of Philosophy* 73, 116–127.

Wikforss, Å. (2008). Semantic externalism and psychological externalism. *Philosophy Compass*, 3, 158–181.

Williamson, Timothy (2006). Can cognition be factorised into internal and external components? In R. Stainton, (ed.), *Contemporary Debates in Cognitive Science.* Oxford: Blackwell, 291–306.