

# Theories of Reference: What Was the Question?

*Panu Raatikainen*

Philosophy, University of Tampere

e-mail: panu.raatikainen@staff.uta.fi

## **Abstract:**

The new theory of reference has won popularity. However, a number of noted philosophers have also attempted to reply to the critical arguments of Kripke and others, and aimed to vindicate the description theory of reference. Such responses are often based on ingenious novel kinds of descriptions, such as rigidified descriptions, causal descriptions, and metalinguistic descriptions. This prolonged debate raises the doubt whether different parties really have any shared understanding of what the central question of the philosophical theory of reference is: what is the main question to which descriptivism and the causal-historical theory have presented competing answers. One aim of the paper is to clarify this issue. The most influential objections to the new theory of reference are critically reviewed. Special attention is also paid to certain important later advances in the new theory of reference, due to Devitt and others.

## **1. Introduction**

In the beginning of the 1970s, the philosophical community experienced a genuine revolution in the philosophy of language – one that had ramifications in many other areas of philosophy. Namely, Saul Kripke (1971, 1972) and Keith Donnellan (1970) famously attacked what was then the prevailing view on the meaning and reference of names, the description theory of reference (or, more briefly, “descriptivism”). Hilary Putnam (1970, 1973, 1975a, b) argued for related points in the case of kind terms (as did Kripke).

The new theory of reference (in short: NTR), as the general view that emerged is often called, has risen to favor. However, a number of noted philosophers have also attempted to reply to the critical arguments of Kripke and others, aiming to vindicate descriptivism. Such responses are often based on ingenious, novel kinds of descriptions, such as rigidified descriptions, causal descriptions, and metalinguistic descriptions. Many seem to be confident that the critical arguments against descriptivism have been neutralized for good.

This prolonged debate raises doubt as to whether various parties really have any shared understanding of what the central question of the philosophical theory of reference is: what is the main question to which descriptivism and NTR have presented competing answers? And more generally: exactly *what* are these theories supposed to be theories of?

One aim of the present paper is to clarify this issue. The most influential objections to NTR are critically reviewed. In addition, the essential content of NTR is elucidated, and the most important developments (in the author's opinion) of NTR are reviewed.

## 2. A Brief Look at the Development of NTR

### 2.1 *Descriptivism and Its Critique*

Though the story should be familiar, let us review the main developments for later reference and in order to fix some terminology. Descriptivism is, according to the definition<sup>1</sup>, the view that the *meaning* of a name is expressed<sup>2</sup>, and the *reference* of the name determined, by the description (or cluster of descriptions) a language user analytically associates with it.<sup>3</sup> Thus, with the name "Socrates", for example, one might associate the description "the Greek philosopher who drank hemlock" and so on.<sup>4</sup> The idea is that such an associated description expresses more explicitly the meaning of a name for the speaker and determines which object is denoted by the name.

---

<sup>1</sup> Before Donnellan and Kripke, there was no explicit school that would have identified itself as "descriptivists" (or advocates of "the description theory of reference"). Rather, Kripke, and to some extent also Donnellan, isolated and abstracted the idea from the rambling literature. Therefore, it was largely up to them to define the view they then critically scrutinized. Given the number of indignant reactions they received, it seems that they were not criticizing a straw man.

<sup>2</sup> Let us sort out one misunderstanding I have sometimes met: the suggested interpretation is emphatically not that the relevant description itself – a linguistic entity – is the sense, or the meaning, of the name. That would obviously be quite an absurd view. Rather, the idea is that the meaning is some sort of abstract or mental entity – perhaps a combination of properties, attributes or concepts – that is expressed more transparently by the associated description, which the referent uniquely satisfies: the referent has exactly those properties (or most of them).

<sup>3</sup> There is also a broader view than descriptivism that Devitt and Sterelny (1999, 62) call "the identification theory": it allows that a speaker may not be able to describe the bearer, provided that he can recognize her; the speaker can, so to say, pick the bearer out in a lineup. This is an improvement, but does not help with names of temporally or spatially distant bearers, such as Cicero or Feynman: with all such names, the descriptivist account is (in this view) the only one possible. And the problems with ignorance and error (see below) remain relevant.

<sup>4</sup> Or, according to the cluster theory version, a larger cluster of such descriptions.

It is very natural to interpret descriptivism also as a theory of *understanding*: to understand an expression is to know its meaning, and that meaning is expressed by the appropriate description or cluster of descriptions (cf. Devitt & Sterelny 1999, 46). Consequently, only by knowing the appropriate description (or cluster of descriptions) and its (their) association with the expression – that is, only by knowing expression’s meaning – can the language user understand the expression and successfully use it in reference; otherwise not.

This view – descriptivism, that is – has its roots in certain remarks by Frege and Russell, although it is a bit unclear whether either of these historical figures actually advocated for exactly the description theory of reference as later defined.<sup>5</sup> Such exegetical questions are, however, orthogonal to the main issue, as there have subsequently been numerous

philosophers who have advocated descriptivism, and in doing so have taken themselves to be followers of Frege or Russell.<sup>6</sup>

In the 1950s, the simple version of descriptivism was increasingly criticized, and – apparently inspired by certain remarks by the later Wittgenstein<sup>7</sup> – John Searle (1958) and Peter Strawson (1959), for example, suggested that the meaning of a name is not expressed by any single definite description, but rather by a *cluster* of descriptions associated somehow more loosely with the name by the speaker. The idea was that a name refers to an entity, which satisfies sufficiently many of these descriptions, though perhaps not all: there is room for some error. The view became quickly predominant. Searle was invited to write an entry on proper names for the most eminent encyclopedia of philosophy at the time

---

<sup>5</sup> Dummett 1973 (110), McDowell 1977, Burge 1979, Evans 1985, Noonan 2001, and Heck & May 2006, for example, have argued against interpreting Frege as a descriptivist. (Currie (1982, 170), on the other hand, defends the descriptivist interpretation.) However, the interpretation of Frege many of these critics favor is more or less the same as “the identification theory” (see note 3 above), which clearly does not save Frege’s view.

<sup>6</sup> In *Naming and Necessity*, Kripke called the simple version of descriptivism “the Frege-Russell view.” Devitt and Sterelny (1999, 45) write, more cautiously, of the classical description theory as “derived from the works of Gottlob Frege and Bertrand Russell.”

In his 2008 Schick prize lecture, Kripke reflects on the issue as follows: “[C]ertainly Frege, like Russell, had generally been understood in this way. This made it important for me to rebut the theory, whether historically it was Frege’s theory or not” (Kripke 2008, 208).

Moreover, (in Kripke 1979, 271, endnote 3), Kripke wrote: “In any event, the philosophical community has generally understood Fregean senses in terms of descriptions, and we deal with it under this usual understanding. For present purposes, this is more important than detailed historical issues.”

<sup>7</sup> Although neither Searle nor Strawson explicitly mentioned Wittgenstein, it is plausible to assume that *Philosophical Investigations*, published in 1953, and the remarks on “Moses” in particular (§79), inspired this view: Kripke suggested the connection in (1980, 31). It is unclear, though, whether Wittgenstein really intended to present any sort of general theory of names, and what exactly was his true aim here; see Travis 1989 and Bridges 2010.

(Searle 1967). This later, more sophisticated variant of the description theory is now commonly called “the cluster theory.”<sup>8</sup> This was the received view in philosophy when Kripke and Donnellan came forward with their critiques.

To begin with, Kripke – already a leading figure in modal logic – presented against descriptivism different arguments based on modal considerations. First, he suggested that ordinary proper names are what he called “rigid designators” (i.e., they refer to the same entity in every possible world<sup>9</sup>), whereas customary descriptions are not; the referent of a typical description (such as “the Greek philosopher who drank hemlock”) varies from one possible world to another; therefore, names cannot be synonymous with such descriptions. Second, Kripke argued that descriptivism entails various “unwanted necessities”: for example, it is not analytically true or necessary that Socrates drank hemlock; it seems possible that events could have gone quite differently. Sometimes one further isolates as a distinct argument “the epistemological argument”: traditionally it has been thought that analytical truths can be known *a priori*; that Socrates drank hemlock is not, on the other hand, knowable *a priori*.

Third, in addition to the modal arguments, both Kripke and Donnellan presented, against both traditional and cluster versions of descriptivism, various “arguments from ignorance and error.” Their central idea was to underline the fact that users of language often have much less knowledge than descriptivism presupposes, or that even the few relevant beliefs they have about the bearer of a name may be false. All a person in the street might be able to say, when asked who Cicero was, is that he was some Roman, or perhaps he could associate with the name “Feynman” only a description such as “some physicist.” Such descriptions are often the most that a speaker can say about the bearer of a name. They are, however, much too loose and general for them to be able to determine the referent of a name; a great number of other people also satisfy such descriptions. Thus people often are much more ignorant than descriptivism requires – and more fallible: a speaker may associate with the name “Einstein” the description “the inventor of the atomic bomb,” or with the name “Columbus” the description “the first European in America” (and no other non-trivial description). Yet the latter description actually picks out some unknown Viking who found America many centuries before Columbus; the former description may not uniquely pick out anyone (the invention of the atomic bomb was group work) or, at best, fit Robert Oppenheimer (who was leading the team), but certainly not Einstein. Nonetheless, it is plausible that even such ignorant and erring language users can successfully use such names to refer to their actual bearers: i.e., to Einstein and Columbus. (If such a sparse or mistaken description is all that a language user can provide, the cluster theory is in trouble

---

<sup>8</sup> My understanding of the situation in philosophy before the revolution has benefited from writings of and personal correspondence with John Burgess.

<sup>9</sup> More exactly, the expression refers to the same entity in every possible world *in which that entity exists*.

just as much.) As Devitt has been fond of saying, descriptivism puts “far too large an epistemic burden” on language users.

As to kind terms, Putnam (1975a) presented his famous Twin Earth thought experiment, inviting us to imagine that there was a planet very much like Earth called “Twin Earth.” We could even assume that each one of us had a *doppelgänger* there and that languages similar to ours were spoken there. There was, however, one peculiar difference: the liquid called “water” on Twin Earth was not H<sub>2</sub>O, but a totally different liquid whose chemical formula was very long and complicated. We could abbreviate it as XYZ. It was assumed that it was indistinguishable from water under normal circumstances: it tasted like water and quenched thirst like water; the lakes and seas of Twin Earth contained XYZ; it rained XYZ there; and so on. Putnam next asked us to roll back time to, say, 1750, when chemistry had not yet been developed on either Earth or Twin Earth. At that time, no one would have been able to differentiate between XYZ and H<sub>2</sub>O. Now Oscar, on Earth, and his *doppelgänger* associate, by stipulation, exactly the same qualitative description to “water.” However, Putnam posited that the extension of “water” was just as much H<sub>2</sub>O on Earth, and the extension of “water” was just as much XYZ on Twin Earth. Putnam’s argument can be viewed as a powerful argument from ignorance.

Kripke noted that description theories can, at least in principle, also be viewed as mere *theories of reference* and not as *theories of meaning*: as such, they only contend that the reference of an expression is determined by the description associated with it (Kripke 1980, 31–32). Kripke went on to argue that (because of the arguments from ignorance and error) at least the familiar forms of descriptivism (involving “famous deeds”) fail also if interpreted as mere theories of reference. However, as Devitt (1981, 13) states, “Description theories are mostly offered as theories of meaning of a name.” Kripke himself added that “some of the attractiveness of the theory [descriptivism] is lost if it isn’t supposed to give the meaning of the name”; this is because it is not clear it can still solve Frege’s puzzles (see below) (Kripke 1980, 33). All in all, it is descriptivism understood as a theory of meaning that is a well-motivated, natural and unified whole, as well as the main target of NTR.<sup>10</sup>

## ***2.2 The Historical Chain Picture***

Kripke also presented a brief sketch of an alternative positive account of reference, the

---

<sup>10</sup> Note that many of Kripke’s key critical arguments (the arguments from rigidity and from unwanted necessity and “the epistemological argument”) only make sense if descriptivism is understood as a theory of meaning. This also suggests that Kripke himself primarily thought of descriptivism in this way.

historical chain picture of reference, or “the causal theory of reference”.<sup>11</sup> Kripke’s picture falls into two parts: there is the initial introduction of a name, and the subsequent transmission of the name or “reference borrowing”.

**Baptism** First, there is the introduction of a referring expression<sup>12</sup> to the language, a *baptism* or a *dubbing event* in which the reference of the name is initially fixed. There, an object must obviously somehow be singled out for naming. According to Kripke, this can happen either with the help of an ostension (by pointing to it or exhibiting it) or of a description. Kripke even adds, “The case of baptism by ostension can perhaps be subsumed under the description concept also. Thus the primary applicability of the description theory is that of initial baptism” (Kripke 1980, 96, fn 42).

Devitt in particular wanted his theory of reference to be more thoroughly causal, also at the stage of introduction of names. He emphasized that, in a *typical* name introduction, those who are present perceive the naming ceremony and the object to be named; by virtue of being in an appropriate causal interaction with each other and the object in the event, they gain the ability to later refer to the object (Devitt 1981, 27). No definite descriptions are needed, and, even if one is involved in the baptism, the named entity may still fail to satisfy it (referential and not attributive use of a description, in Donnellan’s sense). However, even Devitt never denies that names can also be introduced with the help of descriptions, without any perceptual contact or direct causal connection (he simply calls such naming ceremonies “abnormal”).

The baptism need not be purely causal or purely descriptive: it typically involves a categorial concept such as *human*, *place*, or *animal*, but the causal-perceptual part does most of the work. In any case, it was never part of NTR that baptism should always be purely causal. Thus Devitt and Sterelny concede that “the introducer of a name must use some general categorial term such as ‘animal’ or ‘material object’ (Devitt & Sterelny 1987, 65) and “[i]t seems that our causal theory of names cannot be a ‘pure-causal theory’. It must be a ‘descriptive-causal’ theory” (ibid.) Allowing such descriptive elements does not compromise the essence of NTR, as the unspecific descriptive content in question alone is clearly insufficient to determine reference.

Even if a description is essentially used in baptism, it is important to note that the description then normally is not and cannot be

---

<sup>11</sup> The label “causal theory” can mislead and has misled. Even competent philosophers repeatedly interpret the causal theory as claiming that the referent is whatever causes the particular utterance of the name. That is emphatically not the idea. The cause of my utterance of, say, “Aristotle” may be, for example, my friend’s question; it is typically not Aristotle himself.

<sup>12</sup> That is, introduction as the expression with *this specific reference*. It is obviously possible and even common that the (syntactically) same name has already been used in the community with a different reference.

- (i) a description in terms of “famous deeds” (e.g., “taught Plato,” “drank hemlock,” etc.) characteristic of pre-Kripkean descriptivism<sup>13</sup> (for the baby has not yet done any of these things).
- (ii) a meta-linguistic or causal description (see below), such as “the thing to which ‘*Titanic*’ refers”, popular among some recent descriptivists (for the name does not yet refer to anything).

Consequently, it does not provide the sort of description that well-developed forms of descriptivism typically utilize (cf. Burgess, 2013, 28–29). Or, as Kripke himself put it:

Two things should be emphasized concerning the case of introducing a name via a description in an initial baptism. First, the description used is *not synonymous* with the name it introduces but rather fixes its reference. *Here we differ from the usual description theorists.* Second, *most cases of initial baptism are far* from those which originally inspired the description theory. *Usually* a baptizer is *acquainted* in some sense with the object he names and is able to name it *ostensively*. (Kripke 1980, 96, fn 42; my emphasis)

**Reference Borrowing** The second, very important part of Kripke’s picture is the idea of “reference borrowing”. Other language users not present at the name-giving occasion acquire the name and the ability to refer with it from those in attendance at the baptism, still others from the former users, and so on. Later users of the expression need not know or be able to identify the referent. It is sufficient for successfully referring that they are part of an adequate “historical” or “causal” chain of language users which goes back to the first users. Speakers may also be largely ignorant of this chain or even from whom they got the name. Even if the expression was originally introduced in the short term by means of a description, that particular description is not usually transmitted with the expression. Nor is any other uniquely identifying description. Nevertheless, it appears that these later users can use the expression to refer successfully.

Devitt (2006, 2008) attempted to develop Kripke’s sparse and sketchy remarks about reference borrowing into a somewhat more systematic theory: Reference borrowing takes place when a name is used in a communication situation. Devitt grants, however, that a mere causal connection is not sufficient for the hearer to borrow the reference from the speaker. The borrowing has to be an intentional act. First, the borrower must have a sufficient level of linguistic sophistication (a rock or a worm cannot borrow a reference). Second, she must understand what is going on, e.g., that the string of sounds (or symbols) is being used as a proper name.

Devitt emphasizes a distinction that is insufficiently clear in the literature on reference, a distinction between what is required at *the initial time of borrowing* the reference of a name and what is required *at the later time of using* the borrowed name. In NTR, the

---

<sup>13</sup> In Searle’s words, what speakers “regard as essential and established facts” about the bearer.

hearer borrowing the reference of a name from a speaker must, *at the time of borrowing*, intend to use it with the same reference as the speaker.<sup>14</sup>

When the name is ‘passed from link to link’, the receiver of the name must, I think, intend when he learns it to use it with the same reference as the man from whom he heard it. If I hear the name ‘Napoleon’ and decide it would be a nice name for my pet aardvark, I do not satisfy this condition (Kripke 1980, 96).

The idea is definitely not, as some have understood it, that the person who has borrowed the reference of a name must, *at the time of later using it*, intend to refer to the same object as the person from whom he borrowed the name. Both the initial borrowing and the later use are intentional actions, but, according to NTR, subsequent use need not involve any intention to defer to the earlier borrowing; it need not involve any “backward-looking” intention (Devitt 2006, 101–102). It may be that some philosophers’ use of the word “deferring” instead of “borrowing” has contributed to this confusion. (Searle (1983, 244), for example, seems to contribute to this confusion.)

I believe that this is exactly where one crucial difference, often not sufficiently recognized, between descriptivism and NTR lies: the spirit of descriptivism seems to require that there must be some sort of description (perhaps one about borrowing) present every time the expression is used. NTR denies this.

Although a typical name introduction event does involve a causal interaction with the bearer of the name, the *causal chain* is, in NTR, primarily a chain of communication between the earlier and later *uses* and *users* of the *name*, and mostly concerns borrowing. It does not require the bearer of the name to be a causal *relatum*. This is clear once we note that NTR has always permitted the introduction of a name through a description, even in absence of the (postulated) bearer. Even Devitt states that “[t]he central idea of a causal theory was that present *uses* of a name are causally linked to first *uses*” (Devitt 1981, 28; my emphasis). It is emphatically not part of NTR to require, as a common misinterpretation suggests, that there must necessarily and always be a (direct) causal relationship between the bearer and the introducers of the name. Or, if the label “causal theory of reference” is reserved for the thoroughly causal picture in which also baptism is causal (à la Devitt), then it was never a claim in NTR (or by Devitt) that the causal theory truthfully applies to all names (and other referring expressions).

---

<sup>14</sup> I am myself inclined, at least tentatively, to go even further: I do not think there has to be any specific intention to use that particular word with the same reference, even at the time of borrowing: perhaps all that is needed is the absence of an actual decision to begin using that name in a new way (like “Napoleon” for a pet), and a rough understanding of how proper names generally function. However, I should emphasize that this is my own personal view and not something that Kripke or Devitt, for example, would clearly state.



### 2.3 *The Varieties of Reference*

Kripke did not even pretend to have presented a well-developed theory: he only briefly sketched what he called an alternative “picture.” “I want to present just a better picture than the picture presented by the received views” (Kripke 1980, 93). Others have attempted to develop more systematic theory on this basis. Devitt’s book *Designation* (1981) in particular has been an important contribution to this effort.

Some critics, e.g., Unger (1983) and Searle (1983, 239), seem to suggest that refuting counterexamples exist for NTR. 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 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”<sup>15</sup>, Putnam “vixen”, and Devitt and Sterelny “pediatrician”. 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.

More exactly, as we have seen, descriptions sometimes play an essential role in the introduction of an expression. Even in that case, reference borrowing does not require the description to be transmitted with the expression. Consequently, it is not necessary for a subsequent language user to know what the description is in order to successfully make a reference using the expression. In some cases, only the experts know the relevant description, and in other cases, the description is lost in history, and nobody presently knows it. In a few cases, such as “bachelor”, nearly all users are aware of the relevant description. But this seems to be more a sociological fact than a necessary requirement for successfully referring using the expression. My own view is that the reference of even this kind of expression can be borrowed without knowing the relevant description.

In his introduction to his influential 1977 collection *Meaning, Necessity, and Natural Kinds*<sup>16</sup>, Schwartz emphasized that one need not assume that all of language operates in just one way. “[T]he [causal] theory is correct about natural kind terms and the traditional theory is correct about nominal kind terms. It is only the belief in the universal application of one view that excludes the other” (Schwartz 1977, 41).

I think we can easily distinguish more than two different categories of (referring)

---

<sup>15</sup> Kripke adds, “But in many or most cases, I think the thesis [descriptivism] is false” (Kripke 1980, 80).

<sup>16</sup> In which, by the way, Schwartz also introduced the now-common label “new theory of reference.”

expressions: Before the heyday of NTR, early Putnam (1962) suggested, based on certain ideas of Quine, that some terms in science are what he called “law-cluster terms.” Namely, Quine (1960, 1963) argued that, even if a word is originally introduced into science by means of an explicit definition, definitions in science are “episodic”: that is, the status of the resulting equivalence of the new word and its “definiens” need not be an eternally privileged status, a necessary truth, or true by convention. Putnam went further by introducing his notion of “law-cluster concepts,” concepts that are implicated in a number of scientific laws. If any of these laws is treated as a necessary condition for the meaning, one is, Putnam submits, in trouble. Putnam’s proposal was not very different from the cluster version of descriptivism.

Early and late in his career, Putnam did not claim that there were no analytic truths or that “vixen”, for example, could not be correctly analyzed as “female fox”. The point is, rather, that words such as “vixen” or “bachelor” are quite rare and special (in Putnam terminology, “one-criterion concepts”) and not representative, nor do they offer a good model for a general theory of meaning. Many other words have no such standing definitions, if the argument is sound.

Now Putnam’s more mature NTR view implies that his earlier law-cluster account must be wrong with respect to natural kind terms (see especially Putnam 1975b, 281). This has apparently led many to think that Putnam abandoned the whole idea. But this is not the case: if we read Putnam’s later work (see, e.g., Putnam 1986; Putnam 1988, 8–11; cf. Putnam 1973, 206), it is clear that he still thought that the general idea of law-cluster concepts was valid; it holds true for some concepts in science (“momentum” may be an example), even if it does not for common natural kind concepts. And I, for one, think that these observations of Quine and Putnam are still worthwhile.

Though Putnam never formulated the issue this explicitly, I think we can look at his works and systematize the following picture: it is plausible that referring expressions (and terms in science in particular) fall into (at least) four different types:<sup>17</sup>

1. One-criterion words, e.g., “vixen,” “bachelor,” and “Jack the Ripper.” (I am inclined to think that “Vulcan” and perhaps also “Phlogiston” belong to this category.)
2. Law-cluster terms: “momentum” may be an example.
3. Observable or manifest natural kind terms – e.g., “gold,” “water,” and “tiger,”<sup>18</sup> – and common proper names, e.g., “Aristotle,” “Mount Blanc,” and “Cologne”.

---

<sup>17</sup> I once put this sort of division forward in a discussion with Devitt, and he more or less agreed.

<sup>18</sup> These terms are first tentatively identified with the help of their observable properties; but it is part of the idea that their extension is determined by their “inner structure,” lineage, or something else, i.e., more theoretical traits that go beyond direct observation (see 2.4). More theoretical natural kind terms may,

4. Observational terms<sup>19</sup>, e.g., “yellow,” “liquid,” and “sour”.

Much confusion has resulted from the assumption that all referring expressions should refer in the one and same way, and that NTR in particular assumes that. But it is plausible that the causal theory of reference applies most smoothly to the third category (although I believe that the reference of any sort of term can be borrowed).

### ***2.4 The Qua Problem***

The causal theory of reference as applied to general terms has been often criticized as follows: Papineau (1979), Dupre (1981), Crane (1991), Segal (2000), and many others have complained that a sample will usually be a member of many kinds.<sup>20</sup> 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 NTR seems to suggest, how can one rule out incorrect kinds of generalizations? This is the so-called *qua* problem (see Devitt & Sterelny 1999, 72–75).

In fact, however, it has long been recognized among advocates of NTR 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). Recall that Devitt and Sterelny granted that some categorial description may be used even in the case of proper names, which may in part rule out the wrong sort of generalizations.

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), 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.

---

rather, belong to the first two categories. Devitt (1981), for example, explicitly mentions “observational natural kinds” in this connection. More recently, some philosophers of language have begun using the term “manifest natural kinds” (e.g. Soames).

<sup>19</sup> I am well aware of the problems with the notion of “observational terms”, but I think we can use a rough and relative notion here in contrast to the other categories (1–3).

<sup>20</sup> Wittgenstein’s critique of ostensive definitions in *Philosophical Investigations* can perhaps be viewed as a predecessor of this critical argument.

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

In such a situation, it may 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 NTR), and there may be some room for conventional choice, but this is not relevant to the fundamental issue. Superficially similar but internally *radically* different objects or substances (as XYZ in Putnam’s argument, for example, is stipulated to be) simply do not belong to the extension, and this is sufficient for the argument in favor of NTR.

## ***2.5 Can Reference Never Change?***

Soon after Kripke’s groundbreaking contribution, Evans (1973) raised a concern: the picture that Kripke had sketched, the initial simple version of the causal theory of reference, apparently entails that the reference of a name can never change: the initial dubbing or baptism fixes it for good. However, it appears that in reality the reference does sometimes change. For example (this is Evans’ example), apparently “Madagascar” was originally used as a name of part of the African mainland. Due to some confusion, it is now used to refer to a large island.

Perhaps the most important further refinement of the causal-historical theory of reference is Devitt’s idea of “multiple grounding.” Devitt has suggested that it is not only the initial dubbing or baptism that determines the reference: a name typically becomes multiply grounded in its bearer in other uses of the word relevantly similar to a dubbing. In other words, other uses 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).<sup>21</sup> This more sophisticated framework allows reference change and makes it possible to explain it.

Unger (1983) has devised some ingenious variations on Putnam’s Twin Earth thought experiment that seem to support contrary intuition.<sup>22</sup> Many of them, and arguably the most

---

<sup>21</sup> In fact, Devitt first proposed this modification in 1974. Putnam (2001) comments on it approvingly. Also, Kripke (1980, 163) acknowledges the need for some such refinement, but does not explicitly show any awareness of Devitt’s specific suggestion.

<sup>22</sup> Also Bach, for example, refers to them; see Bach 1987, 276–277; cf. Bach 1998.

puzzling ones, are based on some radical but unnoticed change in the environment. For example, imagine that all the water (i.e., H<sub>2</sub>O) on Earth was replaced overnight (say, secretly by aliens) with XYZ; what would the extension of “water” here on Earth be after, say, 100 years?<sup>23</sup> The positive picture 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<sub>2</sub>O. But intuitively, this does not seem at all clear: it seems that “water” would sooner or later switch its reference to XYZ (such scenarios are sometimes called “slow switching” cases in the literature). I contend, however, that Devitt’s improvement – the idea of multiple grounding – enables one to reply not only to Evans’ initial concern over reference change, but also to most of Unger’s much-cited alleged “counterexamples” to NTR. Critics of NTR, however, tend to ignore this important development.

I think that Devitt’s idea has a further application that is not often recognized: Consider a name that is initially introduced via description, without any perceptual contact with the bearer. “Neptune” is a plausible candidate (or “the Boston Strangler”). The idea of multiple grounding makes it possible that, if we later come into perceptual contact with the bearer, the name becomes also non-descriptively grounded in it. In Devitt’s terminology, the status of the name can then change from attributive to referential (cf. Devitt 1981, 57). I think something like this may frequently happen with many terms in science.

### 3. What Was the Question?

#### 3.1 *The “Main Problem” of the Theory of Reference*

But what is the essential problem at issue in the theory of reference? Descriptivism and the historical chain picture have been competing answers, but what exactly was the question? What should an adequate theory of reference be able to do?

It is illuminating to take a look at what Searle, a leading figure (perhaps *the* leading figure) of the descriptivist camp before the critical attack of Kripke and others, has had to say in retrospect:<sup>24</sup>

You will not understand the descriptivist theories unless you understand the view they were originally opposed to. At the time I wrote ‘Proper names’ in 1955 there were three standard views of names in the philosophical literature: Mill’s view that names have no connotation at all but simply a denotation, Frege’s view that the *meaning* of a name is given by a single associated definite description, and what might be called the standard logic textbook view that the *meaning* of a name “*N*” is simply ‘called “*N*’’. (Searle 1983, 242; my emphasis)

---

<sup>23</sup> This is my own example, not Unger’s, but I think that it captures fairly the basic idea of many of his cases.

<sup>24</sup> Note that this was written some time *after* the critique by Kripke, Donnellan and others was published.

Searle continues:

Now the first and third of these views seem to be obviously inadequate. If the problem of a theory of proper names is to answer the question, ‘*In virtue of what* does the speaker in the utterance of a name succeed in referring to a particular object?’, then Mill’s account is simply a refusal to answer the question ... But the third answer is also defective. (Ibid.)

In his 1967 encyclopedia entry, Searle had expressed the issue with respect to a particular name “Aristotle”: “The original set of statements about Aristotle [what speakers regard as essential and established facts about him] constitute the *descriptive backing* of the name *in virtue of which* and *only in virtue of which* we can teach and use the name” (Searle 1967). And in his 1971 “Introduction,” Searle discusses what he calls Frege’s “most important single discovery”: “in addition to the name and the object it refers to, viz. its reference, there is a third element, its *sense* (or as we might prefer to say in English: the *meaning* or descriptive content) of the name *in virtue of which* and *only in virtue of which* it refers to its reference” (Searle 1971, 2; my emphasis).

Searle thus clearly thinks that a central task of a theory of reference is to explain *in virtue of what* an expression refers to the entity it actually refers to. These passages also evidently show that, at least to Searle, the question in descriptivism was essentially about the *meaning* of a name, and not just about the fixation of reference – and that he equated Fregean sense, descriptive content, and linguistic meaning.<sup>25</sup> Finally, “the descriptive backing” for Searle is clearly supposed to be non-trivial, and a language user may well fail to have one and consequently fail to refer with the name.

In the opposite NTR camp, Devitt has expressed the same general idea: “The main problem in giving the semantics of proper names is that of *explaining* the nature of the link between name and object *in virtue of which* the former designates the latter” (Devitt 1974).<sup>26</sup> And similar formulations have been common in the literature. Thus Marga Reimer, in her entry “Reference” in *The Stanford Encyclopedia of Philosophy*, states that of the three central issues, the central questions concerning reference, the first is: “What is the *mechanism* of reference? In other words, *in virtue of what* does a word (of the referring sort) attach to a particular object/individual?”<sup>27</sup>

---

<sup>25</sup> This is clear, if not in “Proper names” (Searle 1958), at least in Searle’s 1967 encyclopedia entry.

<sup>26</sup> Devitt expresses the idea in virtually the same words in *Designation* (Devitt 1981, 6). In his much later encyclopedia entry, he writes: “The central question about reference is: *In virtue of what* does a term have its reference? Answering this requires a theory that *explains* the term’s relation to its referent” (Devitt 1998; my emphasis).

<sup>27</sup> The second is, “What is the relation between reference and *meaning*?”, and the third, “What is the relation between reference and *truth*?”

Accordingly, William Lycan, in his recent survey on the theories of reference (Lycan 2006), proposes that the two central questions of the theory of names are:

(1) *The Referring Question*: In virtue of what does a proper name designate or refer to its bearer?

(2) *The Meaning Question*: What and how does a name mean or signify? What does it contribute to the meaning of a sentence in which it occurs?

I contend that we are justified to conclude that the fundamental question of the theory of reference is the following.

**Main question:** In virtue of what does a referring expression refer to whatever it in fact refers to?

Consequently, it is reasonable to require that *any* satisfactory theory of reference should at least answer this question. Indeed, the pre-Kripkean versions of the description theory of reference (both in its simple and more sophisticated cluster forms) do exactly that: according to these versions, a name refers to the entity it refers to because that entity satisfies the description (or the majority of the descriptions in the cluster) that the language user associates with the name. The historical-chain picture, or the causal theory, in turn answers the question with the suggestion that a name refers to its bearer because, roughly, the user of the name stands in an appropriate causal-historical relation to the first uses of the name.

### 3.2 *The Millian View and Frege's Puzzles*

It has been common in both camps, the descriptivists and the advocates of NTR<sup>28</sup>, to take as their point of departure the so-called Millian view of meaning, or “the direct reference theory” (DRT), and its alleged problems (this view is commonly ascribed to John Stuart Mill – hence the name; but the exegetical issue is again less important).<sup>29</sup> By this, one means the simple view according to which (at least in the case of proper names) the *meaning* of a name is simply its referent – the entity it denotes. In the case of general

---

<sup>28</sup> Accordingly, Searle begins his encyclopedia entry (Searle 1967) as well as his 1971 “Introduction” in this way. Kripke also discusses it at the beginning of his first lecture in *Naming and Necessity* (1980, 26–27); also Devitt begins *Designation* (1981, 3–6) by reflecting on the Millian view and its apparent problems. Braun (2006) begins his handbook chapter similarly, and both Reimer (2009) and Lycan (2008, ch. 3) motivate descriptivism in this same way.

<sup>29</sup> It is now quite popular to assume that the second camp, those who favor NTR, advocate the Millian view. Although some do, this assumption is generally the result of confusion, as we shall see.

terms, the analogous view – that the meaning of a general term is just the set of entities that it applies to – has sometimes been called “extensionalism” (cf. Braun 2006).

Today, it is widely thought that such views encounter enormous difficulties in so-called “Frege’s puzzles”.<sup>30</sup> One of them concerns identity statements. Let us consider the example that derives from Frege, the following pair of sentences:

Hesperus is Hesperus.

Hesperus is Phosphorus.

It is now well known that the two names “Hesperus” and “Phosphorus” denote in fact the same heavenly body, namely the planet Venus. However, this was not known in ancient times. Rather, the names were used as if they referred to different heavenly bodies: “Phosphorus” to a bright star visible in the morning, and “Hesperus” to a bright star visible in the evening. Now if the Millian view was right, it should follow<sup>31</sup> that the above two sentences would also have the same meaning. Nevertheless, the first of them is trivial and knowable *a priori*, whereas knowing the latter requires substantial empirical knowledge. It does not appear to be analytically true. Consequently, it is not plausible that the sentences have the same meaning. An analogous problem can be presented, in the case of general terms, for extensionalism, e.g., with the predicates “renate” (creature with a kidney) and “cordate” (creature with a heart) – presumably (at least, so the argument assumes) these have the same extension, but it is difficult to maintain that they have the same meaning.

The Millian view also runs into trouble with names without a referent such as “Father Christmas” or “Vulcan”<sup>32</sup>. It seems to entail that such names have no meaning at all. Apparently, however, sentences containing such names – and accordingly the names themselves – are perfectly meaningful. Consider:

Vulcan is a planet.

---

<sup>30</sup> Whether or not the actual, historical Frege intended his “senses” to be linguistic meanings (or a central aspect of meaning, or something close), his arguments work beautifully with respect to linguistic meaning and have consequently become – thus interpreted – standard in the philosophical theory of meaning (see, e.g., Searle 1967 and 1971; Devitt & Sterelny 1987 and 1991; Braun 2006; Lycan 2008). However, the problem of empty names was in reality more central to Russell than to Frege. Four different puzzles are often mentioned, even, but I must be brief and condense here.

<sup>31</sup> That is, under some plausible assumption of compositionality or the principle of substitutivity of synonyms.

<sup>32</sup> “Vulcan” was the alleged directly unobservable planet scientists once postulated, orbiting between Mercury and the Sun, and causing the deviations in Mercury’s orbit. However, it turned out that there exists no such thing.



Vulcan does not exist.

It seems therefore plausible – pace Millianism – that a name can have meaning even if it does not refer to anything real. At least the latter sentence even seems true.

Consequently, descriptivism, put forward as a more plausible alternative to the Millian view of *meaning*, proposes that there must be more to the *meaning* of a name than the referent, the entity named – namely the descriptive content of the name (what the associated description, or the cluster of descriptions, expresses).

Indeed, descriptivism has been standardly motivated by referring to Frege’s puzzles. It has been frequently considered to be a major virtue of descriptivism, as opposed to the Millian view, that it can so neatly solve the puzzles. This is clearly the case, for example, in Searle’s above-mentioned encyclopedia entry (Searle 1967). Also, another leading contemporary descriptivist, Bach, writes, “Avoiding these puzzles is the theoretical motivation underlying any description theory of names” (Bach 1987, 134). Further, Katz (1990), still another prominent descriptivist, also motivates descriptivism with Frege’s puzzles. Chalmers (2002), too, presents his “broadly Fregean account of meaning” by starting with Frege’s puzzles.<sup>33</sup>

Thus it also appears reasonable to require that any well-motivated form of descriptivism should at least be able to deal with Frege’s puzzles.<sup>34</sup> And if so, descriptivism must also be understood as a theory of meaning (and not merely as a theory of reference-fixing) in the first place.

### 3.3 Shared Meanings

The view that (conventional linguistic) meanings are and must be intersubjectively shareable and public has been widely advocated in the analytic tradition of philosophy.<sup>35</sup> Already Frege held that meaning (or “sense”) is, in general, shareable:<sup>36</sup> mankind has,

---

<sup>33</sup> Chalmers is not a flag-carrying descriptivist, but, rather, distances himself from descriptivism. Nevertheless, he suggests that a meaning of a description can “approximate” the meaning of the original expression (see, e.g., Chalmers 2002, 149, 160). As will become evident, I do not believe that is true.

<sup>34</sup> Everett (2005) makes essentially the same point.

<sup>35</sup> Russell, at least in his “The philosophy of logical atomism” (1918), has been a notable exception: to him, language is essentially private. This view became quite unpopular due to the criticism of the later Wittgenstein.

<sup>36</sup> There may be, for Frege, some exceptions in the case of indexicals. Moreover, Frege grants that different speakers may attach different senses to a name. But for Frege, this was more an unhappy shortcoming of

according to him, “a common store” of meanings (“thoughts”, as he called the senses of sentences). Such a meaning can be expressed in different languages, and is objective. The meaning (sense) of an expression or a sentence is what one grasps in understanding it: it is grasped by everyone sufficiently familiar with the language in which it belongs. (Frege 1892a, b) The latter idea is reflected, e.g., in Carnap: “If we understand the language, then we can grasp the sense of the expressions.” (Carnap 1947, 119).

This sort of view of understanding, which quite literally identifies understanding with knowledge of meaning, has been labelled by Miller “the epistemic conception of understanding”.<sup>37</sup> This picture and the shareability of meaning play an essential role in the philosophy of Dummett and his disciples, for example. Dummett too has been another important recent advocate of descriptivism.<sup>38</sup>

Indeed, it has been quite common, from Frege onwards, to think that successful communication requires shared meanings. Furthermore, the idea that whatever meaning is, it is what the expression and its translation share (i.e. they have “the same meaning”) has been highly influential in the philosophy of language, especially after Quine. All such considerations point towards a largely public community-wide meaning.

In contrast, the idea of radical meaning variance was famously suggested, in the context of the philosophy of science, by Kuhn and Feyerabend in the 1960s: their proposal was that the meaning of an expression occurring in a scientific theory (or in a system of beliefs) changes when the theory is modified or replaced by another theory in which that expression also occurs.<sup>39</sup> This led to their notorious thesis of incommensurability. It was soon recognized that this idea entails conclusions that are highly implausible and even inconsistent with its own starting points (see, e.g., Shapere 1964 and 1966; Achinstein 1968). No doubt the meanings of expressions sometimes change<sup>40</sup>, but it is not reasonable to postulate this massive and extreme meaning variance.

It is therefore plausible to assume that meanings (in the relevant sense) are shared and, by and large, stable, so that two persons can believe and state quite different things about the

---

natural language, something that would be eliminated in the ideal logical language. Furthermore, such a difference of senses amounted to, for Frege, speaking really different languages. Be as it may, one should not one-sidedly exaggerate this aspect of Frege’s views on sense at the expense of just how central the objectivity and the shareability of meaning was for him (see, e.g., May 2006 and Kremer 2010).

<sup>37</sup> “Call the psychological thesis that a speaker’s understanding of a sentence consists in knowledge of its meaning *the epistemic conception of understanding*” (Miller 2006, 994).

<sup>38</sup> Or, strictly speaking, of the more general “identification theory”; see footnotes 3 and 5.

<sup>39</sup> Their “contextual theory of meaning” can be viewed as a version of descriptivism.

<sup>40</sup> Indeed, Devitt’s idea of multiple grounding (see above) provides one account of how this could happen.

same subject matter and contradict each other but still mean the same by their relevant words. More logically, it should be possible for two sets of statements, beliefs or theories, even if quite different, to stand in logical relation to each other: in particular, be inconsistent with each other. However, this requires that the meanings of **the** relevant expressions must be the same. In other words, we often want to have a genuine disagreement and not merely equivocation or talk past each other. In fact, Frege realized this. According to him, if meanings were subjective, a common science would not be possible: “It would be impossible for something one man said to contradict what another man said, because the two would not express the same thought at all, but each his own” (Frege 1914).

Meanings do occasionally change. And some degree of variation of meaning in a larger linguistic community is a fact of life. However, one should not exaggerate the frequency with which this occurs. I contend that it is rational to follow here a maxim that Putnam (1965, 130) has playfully called “Occam’s eraser” (the idea goes back to Ziff):

(OE) Differences of meaning are not to be postulated without necessity.

Descriptivism in particular, when interpreted as a theory of understanding in accordance with the epistemic conception of understanding, presupposes that meanings are public and shared: the idea is that learning and understanding an expression requires the *correct* descriptive content to be associated with the expression (the one that other competent speakers already associate with it); not just any arbitrary subjective association suffices.

In his 1967 encyclopedia entry, Searle described the situation as follows. What speakers (already competent with the expression) consider as *essential* and established facts about the bearer constitutes *the descriptive backing* of the name, and some indefinite subset or the disjunction of these descriptions is *analytically* tied to or *logically* connected with the name. We teach and learn to use the word with that descriptive backing. It is natural to understand all this as meaning that not just any subjectively associated description would do. It is also worth noting that Searle, in his classic paper of the modern description theory, “Proper names”, criticized the more traditional simple versions of descriptivism, among other things, on the ground that they would entail that a name “would have *different meanings* for different people” (Searle 1958, 169). He clearly assumes that this is a defect and that a satisfactory theory should provide a stable meaning shared by different speakers of the community. For his part, Strawson (1969) reflects, for example, how children learn to master the meaning rules of the language through conditioning and training by adult members of the community. The adults teach “the same, the common language” and, as a consequence of this, it is a natural fact that language and linguistic meaning are public.

Emphasizing the public and shared nature of meanings does not, obviously, mean that it would not be possible to identify and scrutinize different, more subjective notions of

“meaning”, such as “coloring” and “shading” (Frege), “expressive meaning” (Carnap), “speaker’s meaning” (Grice), “conception” (Burge) or whatever else. However, the fact remains that the sort of meaning that has been the main focus in the honorable analytic tradition is the public and stable one. And that is what we can plausibly assume is at stake in the debate surrounding descriptivism as well.

Consequently, it is also natural to demand that a satisfactory theory of meaning and reference should provide socially shared public meanings – that it would not entail that the meaning of a referring expression varies wildly from one person to another, even inside a particular linguistic community. Otherwise, it is unclear whether the theory is even a contribution to the same debate.<sup>41</sup>

**Interim conclusion** We can summarize the reflections of the above few subsections as follows. It is plausible and fair to require that any satisfactory theory of reference and meaning – and any well-motivated version of descriptivism in particular – at least

- (i) is able to answer the main question (“In virtue of what does a referring expression refer to whatever it in fact refers to?”);
- (ii) is able to accommodate to Frege’s puzzles; and
- (iii) provides relatively stable and public meanings.

### ***3.4 Meaning, Understanding, and Manifestability***

Let us briefly digress to a different topic. Namely, quite independently of descriptivism, the view that meanings must be not only shared but also exhaustively manifested in the observable (non-linguistic and linguistic) behavior has been quite influential in modern philosophy of language. This has been a common theme in the otherwise different views of such towering figures as Dummett, Quine and Davidson – and their many followers.

It is easiest to begin with Dummett, because there is apparently less controversy over what his view is, and he is most closely connected to the debates we have already discussed: Dummett was a devoted Fregean and advocated for a broadly descriptivist interpretation of “sense”.<sup>42</sup> However, he gives his view an important twist not found in Frege: he adds that

---

<sup>41</sup> Chalmers, for example, recently granted that in his two-dimensional “broadly Fregean” approach, the senses, or intensions, “do not play the ‘public meaning’ role” (Chalmers 2012, 249). Consequently, I think it is somewhat misleading to even put that theory forward as a statement in the mainstream debate on meaning and reference (as Chalmers has repeatedly done). Chalmers, though, justifies all this with vague gestures towards Frege: that he allowed his senses to vary between different speakers as well (ibid., 251). See, however, footnote 36 and the discussion of Frege above.

<sup>42</sup> More precisely, Dummett advocated for “the identification theory” (see footnote 3), but, as was noted above, in a great many cases this makes no difference.

learning language and communication require meanings to be fully manifested in observable behavior. This led Dummett to his semantic antirealism: the view that some sentences, although they have determinate meaning, are neither true nor false (Dummett 1978, 1993).

Quine is famous for his thesis that translation, and accordingly meanings, are indeterminate. There has been less agreement, however, on what exactly his key reasons or premises are for this view. My own interpretation (defended in some detail in Raatikainen 2005) is that the thesis is essentially grounded in his “linguistic behaviorism”<sup>43</sup>, a view not that different from Dummett’s manifestability requirement. Language learning, according to Quine, turns on intersubjectively observable features of human behavior and its environing circumstances, “there being no innate language and no telepathy.”

Finally, even if the admirers of Davidson tend to emphasize the differences between Davidson and Quine, it is hardly controversial that Davidson took for granted linguistic behaviorism he learned from Quine: “Meaning is entirely determined by an observable behavior, even readily observable behavior” (Davidson 1990, 314). To be sure, Davidson then went on to develop his own original program, but that general view formed the basis for it.

The founding fathers of NTR did not much comment on the issue of manifestability<sup>44</sup>, and even in later literature, the relationship between manifestability and NTR is rarely touched upon. Although the tension between NTR and Dummett’s views had not gone unnoticed but had been observed by several philosophers, including Dummett himself,<sup>45</sup> I believe it was Devitt (1983) above all that made this clash current. The connection with Quine had been suggested in the literature in passing a few times, but was properly treated only in my own article (Raatikainen 2005). (I am not aware of any discussion of NTR and Davidson in this respect; but I believe that what I said about Quine largely applies.)

Briefly: Consider, for example, once again Putnam’s Twin Earth story and the year 1750. One simply could not determine, on the basis of observable linguistic behavior of language users in 1750, whether our “water” and “water” on Twin Earth had the same meaning or not. The manifestable uses of the two linguistic communities would be indistinguishable –

---

<sup>43</sup> This view is somewhat different from and more specific than the traditional all-encompassing behaviorism and is not obviously refuted by “the standard objections to behaviorism” (as I argue in Raatikainen 2005; see especially footnote 47).

<sup>44</sup> Putnam (1975c) did comment on Quine’s indeterminacy thesis, but apparently failed to see the possibility that NTR would undermine it. Furthermore, he presented more or less the relevant argument as a critique of Sellars in Putnam 1974, but seemingly did not see its significance for Quine’s thesis.

<sup>45</sup> See Millar (1977), McGinn (1982), Currie & Eggenberger (1983), and Dummett (1974); cf. Raatikainen (2010).

as would be any explicit verbalizable knowledge of meaning.

Nevertheless, under the standard assumption that meaning determines extension — that is, if the extensions of two expressions differ, their meanings cannot be the same — it is the case that our “water” and “water” on Twin Earth do differ in meaning. This thought experiment and its kin undermine the assumption of the manifestability of meaning and the whole equation of competence in a language with knowledge of its meanings, i.e. the epistemic view of understanding.

All this is perfectly natural from the NTR point of view, according to which reference is determined by a historical chain largely opaque to a language user (often even to the entire community within a fixed time period). Inasmuch as difference in reference is sufficient for difference in meaning (a widely shared assumption), there can be differences in meaning that are not detectible on the basis of observable behavior. Hence linguistic behaviorism is false.

## **4. New Forms of Descriptivism**

Let us now revisit various new forms of descriptivism which have been proposed as responses to the critical arguments of Kripke and others.

### ***4.1 Rigidified Descriptions***

As Plantinga (1974) first noted, Kripke’s argument based on rigid designation can be circumvented if, instead of simple descriptions, one focuses on descriptions that have been “rigidified.” For example, one may associate with the name “Socrates” a description of the following type:

The philosopher who *actually* (in the actual world) drank hemlock.

Such rigidified descriptions, just as proper names, refer to the same entity in every possible world. There is no question that this modification to descriptivism can bypass Kripke’s argument from rigid designators. Furthermore, it is altogether possible that the philosopher who actually drank hemlock might not have drunk hemlock. In other words, this version of descriptivism does not lead to unwanted necessities either. These facts have led many eager defenders of descriptivism to think that rigidified descriptions can save the description theory of reference. This is, however, premature; the situation is not quite that bright for descriptivism.

To begin with, it is intuitively implausible that a simple sentence such as “Socrates is snub-nosed” would have the actual world in its entirety as part of the subject matter of what is

said by the sentence (cf. Fitch 2004, 48). Besides, in the context of possible world semantics<sup>46</sup>, descriptivism based on rigidified descriptions faces difficulties. Namely, there are problems with it in the context of Frege’s puzzle concerning identity statements: names with intuitively different meaning but with the same referent now refer to the same entity in all possible worlds, and consequently have – in the possible-worlds framework – the same *intension* (which is the technical explication of the notion of *sense* or *meaning* in this context); but the names were presupposed to have different meanings.<sup>47</sup> (Soames (2001) also presents more complicated technical critique of rigidified descriptions.)

Furthermore, the epistemological argument has not disappeared. If it was not analytically true and *a priori* knowable that Socrates drank hemlock, neither is it analytically true and knowable *a priori* that Socrates is the philosopher who in the actual world drank hemlock. Finally, the weighty arguments from ignorance and error have not been circumvented. If the best description a speaker can provide for, say, “Feynman” is “some physicist,” or a speaker associates with “Einstein” the description “the inventor of the atomic bomb,” rigidifying such insufficient or false descriptions does not provide a way out. In sum, rigidified descriptions cannot save descriptivism.

#### **4.2 Causal Descriptivism**

A popular recent version of the description theory of reference is the so-called “causal descriptivism” favored by David Lewis (1984), Fred Kroon (1987), and Frank Jackson (1998), for example. Its ingenious idea is the suggestion that associated with a name “*N*” is a description, roughly, of the form “the entity which stands in the appropriate causal-historical relation (which accords with the causal theory of reference) to the name.” Slightly more exactly, according to this theory, speakers associate with a name “*N*” a description of the form

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. the causal-historical chain picture) theory of reference.

Its popularity notwithstanding, causal descriptivism has a few serious problems. Devitt and Sterelny (1999, 61) summarize them accurately. First, it is *psychologically implausible*: it requires that every competent speaker must possess a theory of reference – the absolutely

---

<sup>46</sup> Personally, I think that possible world semantics has, in the philosophical theory of meaning, rather limited value. But some recent descriptivists attribute to it a highly central role. The following observation has some bite against such descriptivists.

<sup>47</sup> I have borrowed this observation from Cumming 2016. (The first version appeared in 2008.) I do not know whether it originated with Cumming or whether it has an earlier history, nor do I know who exactly to credit for it.

correct and complete theory of reference –and it is doubtful that anyone possesses such a theory. Second, the descriptions it provides are *parasitic and redundant*: if it is true, it admits that a name stands in a causal-historical relationship, *R*, to its bearer; *R* alone is sufficient to explain reference, and further description involving *R* is redundant.<sup>48</sup>

Finally, if we are interested in the question of the *meaning* of proper names and not only in the fixation of reference, as we should be, causal descriptivism is quite problematic (see section 4.4 below).

### ***4.3 Nominal Descriptivism or Metalinguistic Descriptivism***

A third popular new version of the description theories is so-called “nominal descriptivism” or “metalinguistic descriptivism”.<sup>49</sup> It has been advocated as a response to the arguments of Kripke and others, e.g., by Searle (1983), Bach (1987), and Katz (1990 and 1994), and apparently Chalmers (2002) has built it into his more general “two-dimensional” theory.<sup>50</sup> This approach attempts to circumvent the powerful arguments from ignorance and error with the suggestion that surely even the ignorant and erring language users can associate with a name “*N*” a description of the form:

The thing to which “*N*” refers.

But it is now important to note that this theory does not even begin to answer the main question: In virtue of what does an expression refer to whatever it refers to? The theory already presupposes the reference relation, and cannot explain it.

Searle should be credited, though, for being at least to some extent aware of this problem.<sup>51</sup> So he adds that ignorant speakers can use descriptions of this sort, but there must be other speakers who know some more substantial descriptions to whom the ignorants can “defer” (Searle 1983, 243). Now one problem with this move is that it entails that the name has a different meaning for the ignorant speakers and for the more knowledgeable ones in a

---

<sup>48</sup> Cf. Kripke 1980, 70: “Obviously if the only descriptive senses of names we can think of are of the form ‘the man called such and such’, ‘the man called “Walter Scott”’, ‘the man called “Socrates”’, then whatever this relation of *calling* is is really what determines the reference and not any description like ‘the man called “Socrates”’. Though Kripke’s target here is more nominal or metalinguistic descriptivism (see below), his point seems to be more or less the same.

<sup>49</sup> Sometimes, however, it is counted as a version of causal descriptivism.

<sup>50</sup> Or should Chalmers’ view be classified as a version of causal descriptivism? I am not sure. In any case, his formulation is, “The person called ‘*N*’ by those from whom I acquired the name.”

<sup>51</sup> Recall Searle’s above-cited statement that “the standard logic textbook view” – which is simply metalinguistic descriptivism – is “obviously inadequate”.



single linguistic community – a sort of consequence which Searle has criticized in other contexts, and which is, in any case, ad hoc and implausible.

Nominal or metalinguistic descriptivism also seems to be in conflict with the spirit of descriptivism. Let us recall what Strawson, another key figure of modern descriptivism, said: “[I]t is no good using a name for a particular *unless* one knows who or what is referred to by the use of the name. A name is worthless without a backing of descriptions which can be produced on demand to explain the application” (Strawson 1959, 20). Also: “One cannot significantly use a name to refer to someone or something *unless* one knows who or what it is that one is referring to by that name. One must, in other words, be prepared to substitute a description for the name” (*ibid.*, 181). Searle also seemed to share the same spirit (in the passage we already cited): “The original set of statements about Aristotle [what speakers regard as essential and established facts about him] constitute the descriptive backing of the name in virtue of which and *only* in virtue of which we can teach and use the name” (Searle 1967).

In other words, descriptivism, in its original pre-Kripkean form, required a speaker to have some *non-trivial* identifying knowledge of the bearer of the name – otherwise she fails to successfully use the name to refer to anything. Metalinguistic descriptivism, on the other hand, completely trivializes the issue: a speaker can always, almost trivially, provide a metalinguistic description.

From the perspective of the more traditional descriptivism, metalinguistic descriptivism results in what might be called “miraculous competence”. Imagine a monolingual English-speaker, Jack, who does not understand a word of French. But assume then that Jack associates, with every French proper name ‘*N*’ he faces, a description of the form “the thing to which ‘*N*’ refers” and, analogously, with every French predicate ‘*P*’ a description of the form “the entities which are in the extension of ‘*P*.’” According to metalinguistic descriptivism, Jack should now be also able to successfully refer with all these names and predicates of French. This seems to be totally contrary to the original spirit of the description theory of reference.<sup>52</sup>

Finally, the descriptions that metalinguistic descriptivism provides are even more parasitic and redundant (see above) than those of causal descriptivism.

#### ***4.4 A Theory of Meaning?***

I take it that the gist of descriptivism has always been – in contradistinction to the Millian

---

<sup>52</sup> From the perspective of NTR this may not be unacceptable. My point here is simply to underline just how different this is from the spirit of traditional, pre-Kripkean descriptivism.

view – that the *meaning* of a name is something more than the referent and that the associated description expresses this meaning. Nevertheless, both causal and metalinguistic descriptivism are in fact quite implausible as theories of meaning (and do not even “approximate” meaning, as Chalmers suggests).

Not only can co-referential names be non-synonymous (Frege’s puzzle of identity), but presumably distinct expressions can be synonymous, that is, share the same meaning. For example, apparently the proper names “Köln” (in German) and “Cologne” (in English) have the same meaning.<sup>53</sup> But the descriptions

“the thing to which ‘Köln’ refers”

and

“the thing to which ‘Cologne’ refers”

attach *different* descriptive content, or meanings, to them.<sup>54</sup> Causal descriptivism has, *mutatis mutandis*, similar problems. Consider the Swedish sentence

*Jultomten är vänlig*

and the English sentence with the same meaning,

“Father Christmas is kind.”

However, both causal descriptivism and metalinguistic descriptivism ascribe different meanings to *Jultomten* and “Father Christmas”; consequently, the entire sentences should also have different meanings. This consequence is unnatural and highly implausible.

There are similar problems with general terms. “Woodchuck” and “groundhog” are synonymous. However, their descriptions<sup>55</sup>,

---

<sup>53</sup> As we have noted, Frege explicitly contended that different expressions may well have the same sense: “The same sense has different expressions in different languages or even in the same language” (Frege 1892a, 159). Furthermore, many philosophers take the apparent fact that distinct expressions may be synonymous, that is, have the same meaning, as part of the basic data that any philosophical theory of meaning should be able to explain; see e.g. Lycan 2008, 65–66, 78. Also Chalmers (2002, 139), for example, explicitly grants this possibility.

<sup>54</sup> Such line of argument is by no means original with me. I picked it up from Putnam (1988, 27). It now seems to me that Kripke (1979, 274, endnote 12) is making essentially the same point. I developed the idea already in (Raatikainen 2006). Everett (2005) makes more or less the same observation. The general idea of such translation arguments goes back to Church’s critique of Carnap (Church 1950).

<sup>55</sup> Metalinguistic descriptivism is typically presented explicitly only for singular names; but I assume that if it is supposed to work for general terms too (and recall that there are also Frege’s puzzles for them to be dealt with), it uses descriptions like the ones given here (details are irrelevant for the general point here).

“the entities which are in extension of ‘Woodchuck’”  
and  
“the entities which are in extension of ‘Groundhog,’”

attach different descriptive content (i.e. meaning) to them. Causal descriptivism has the same problem. At worst, some versions<sup>56</sup> of causal and metalinguistic descriptivism imply that even distinct utterances of a single name cannot have the same meaning.

A quite obvious countermove explicitly advocated by Kroon and Jackson is to simply deny that causal or metalinguistic descriptivism is even intended to be a theory of meaning – that it is merely a theory of reference fixation.<sup>57</sup> But it is important to note that, if this line is taken, then these versions of descriptivism cannot even begin to deal with Frege’s puzzles, e.g., explain the difference in *meaning* of co-referential but non-synonymous names. Yet, that has always been a major motivation for descriptivism (see section 3.2 above). Kroon notes this and is ready to bite the bullet. Jackson apparently never even mentions the puzzles. One may feel, with some justice, that too much has now been given up.

Katz in turn contends that, at least in the realm of proper names, two distinct names can never be synonymous. I, for one, cannot help feeling that such a line of response is intolerably ad hoc and ignores many actual cases that appear, at least *prima facie*, to provide counterexamples. Moreover, Katz abandons the key Fregean assumption that meaning determines reference. And this is a defense of the description theory of *reference*?

Such watered-down versions of descriptivism may be a bit more defensible, but they are impotent in addressing the questions the description theory of reference has standardly been developed to provide answers to.

#### ***4.5 Substantial and Trivial Versions of Descriptivism***

It appears to be a common assumption that NTR was aimed at refuting descriptivism in any possible form. Some philosophers then argue that Kripke and others in fact fail to demonstrate this general conclusion: devices such as metalinguistic or causal descriptions are presented as ingenious and effective responses to the critical arguments of Kripke and

---

<sup>56</sup> Often such versions use, more exactly, descriptions such as: “The entity standing in relation *R* to my current use of the name ‘*N*’”; or “The entity called ‘*N*’ by my interlocutors”.

<sup>57</sup> Jackson nevertheless says (1998, 206) that names are *abbreviated* descriptions; it is difficult indeed to understand what this is supposed to mean, if not that names are synonymous with descriptions.

others. However, if one bothers to look at what Kripke and Donnellan actually said in their seminal texts, one sees that they did not declare a complete and unconditional victory. If one reads only these confident contemporary descriptivists, it would be difficult to guess that, in fact, Kripke and Donnellan were well aware of the possibility of something like causal descriptivism or metalinguistic descriptivism.

Kripke asked whether descriptivism could be rescued:

[T]here is a sense in which a description theory must be *trivially true* if any theory of the reference of names, spelled out in terms independent of the notion of reference, is available. For if such a theory gives conditions under which an object is to be the referent of a name, then it *of course* uniquely satisfies these conditions. (Kripke 1980, 88, fn 38; my emphasis)

Kripke called such a theory “a trivial fulfilment” of descriptivism (*ibid.*, 162). In the “addenda” to *Naming and Necessity* (160–62), he also explicitly discussed metalinguistic descriptivism. Kripke noted that “the resulting description would hardly be one of the type which occurs to a speaker when he is asked such a question as, ‘Who is Napoleon?’, *as the description theorists intended*” (*ibid.*, 162; my emphasis). In the footnote quoted above, Kripke wrote: “however, the arguments I have given show that the description must be one of a *completely different sort* from that supposed by Frege, Russell, Searle, Strawson and other advocates of the description theory” (my emphasis).

For his part, Donnellan noted that it is necessary to add some qualifications to “the principle of identifying descriptions” (his label for descriptivism), namely to require that the descriptions provided are “*non-question-begging*” if descriptivism is supposed to be an interesting view at all. This is because

there are certain descriptions that a user of the name (providing he can articulate them) could always provide and which would always denote the referent of the name (providing there is one). *No argument could be devised to show that the referent of a name need not be denoted by these descriptions.* At the same time anyone who subscribes to the principle of identifying descriptions would hardly have these descriptions in mind or want to rely on them in defence of the principle. (Donnellan 1970, 365; my emphasis).

As examples of such “question-begging” descriptions, Donnellan mentions “the entity I had in mind” and “the entity I referred to.” He contends that if descriptions such as these are included in the “backing descriptions”, descriptivism “would become *uninteresting.*” Donnellan then points out that Strawson, for example, explicitly excludes descriptions such as these. In Searle (1958 and 1967), there are also passages that suggest Searle would, likewise, not at the time have accepted such “question-begging” descriptions, either.

In a similar spirit, Putnam (1970) wrote:

In the traditional view, the meaning of, say ‘lemon’, is given by specifying a conjunction of properties. ... *In one sense, this is trivially correct.* If we are allowed to invent unanalyzable properties ad hoc, then we can find a single property – not even a conjunction – the possession of which is a necessary and sufficient condition for being a lemon, or being gold, or whatever. Namely, we just postulate *the property of being a lemon*, or *the property of being gold*, or whatever may be needed. If we require that the properties  $P_1, P_2, \dots, P_n$  not be of this ad hoc character, the situation is very different. (Putnam 1970, 140; my emphasis)

Consequently, it is odd that so many philosophers present causal or metalinguistic descriptivism as conclusive responses to Kripke and others, as the latter were aware of the possibility of such trivial and ad hoc descriptions from the beginning. The argument was always specifically about the well-motivated and non-trivial types of descriptivism that were popular in the literature then. It was never denied that some ad hoc and trivial variants of descriptivism could possibly circumvent the critical arguments. Therefore, it is hardly a great philosophical achievement to now put forward such versions of descriptivism.

**Interim conclusions** It seems fair to conclude that many of the recent defenses of descriptivism amount to moving the goalposts. These new forms of descriptivism

- are often impotent in answering the main question (i.e., in explaining in virtue of what does a name refer to a specific entity).
- provide often only parasitic and redundant descriptions (see 4.2 above).
- often fail to provide a socially shared intersubjective meaning, but make meaning differ wildly between different speakers.
- often cannot solve Frege's puzzles, which was always a key motivation of descriptivism.

## 5. Kind Terms

The literature on general terms has largely focused on Putnam's Twin Earth thought experiment and “water.” Jackson (1998) and Chalmers (2002), for example, suggest in response that the meaning of “water” can, after all, be captured (at least approximately) with a description. If we abbreviate “the clear drinkable liquid that fills the lakes and rivers, falls from the sky in rain (etc.)” as “the watery stuff”, they propose that a description such as “the watery stuff *I am acquainted with*” or “the watery stuff *found around here*” would be sufficient. It picks out  $H_2O$ , but would have referred to XYZ if it had been introduced by someone on Twin Earth, or if the watery stuff on Earth had been XYZ. Whatever the details, this example has some importance for Jackson and Chalmers: it is supposed to provide a paradigm of the foundations of conceptual analysis or the Fregean sense of a kind term.

Although “water” served its purpose in Putnam’s science fiction, it is in many ways an atypical example and can be quite misleading: Almost three-quarters of Earth is covered by water; our bodies are mostly made up of water; and clean drinking water is fundamental to our survival. Therefore, we are all enormously familiar with water. We are in touch with water on a daily basis. Not so with many other natural kinds. By focusing on an atypical and extreme example, Jackson and Chalmers smooth away the relevance of reference borrowing and the phenomenon of ignorance and error so central to NTR.

Consider, instead, the following scenario: Imagine uneducated peasants somewhere in northern Europe some time ago (perhaps even in the Middle Ages). Assume they have picked up some kind words from their priest, who has been reading the Bible to them. They have heard, in *The Song of the Songs*, about gazelles and leopards, cedars, firs and fig trees, spikenard and saffron, myrrh and aloes, pomegranates, sapphires, and alabaster. However, it may be quite unclear to them what these things really are – even whether they are trees, flowers, mammals, predators, metal, or something else.<sup>58</sup>

Nevertheless, if the historical chain picture is at least roughly correct, even such ignorant people can borrow the reference, have the word, and use it to refer successfully. However, there is not much descriptive content in their minds about these kinds to go on – nothing like “the watery stuff” description. For contingent reasons, we all know a great deal about water (even before we know its chemical formula is H<sub>2</sub>O), but knowing that much is neither typical nor necessary for successful reference with a kind term. Moreover, it seems that reference-determining meanings must be something quite different from what is expressed by descriptions in the style of “watery stuff”.

## 6. Back to the Millian View?

One noteworthy, more recent development in the theory of meaning has been the revival of the Millian view. Several philosophers have concluded not only that the critical arguments against descriptivism are decisive, but also that they leave no choice but to return to *the direct reference theory* (DRT): the meaning of a proper name is, after all, simply the object denoted.<sup>59</sup> (Accordingly, many of those who are sympathetic towards descriptivism seem

---

<sup>58</sup> There is a problem with this example: as it is presented, it involves translation from Hebrew to English, and the early translations of the Bible were quite inadequate. A cleaner example would be one with a Jewish community hearing all this in the original Hebrew. I only wanted to present the example (including the words) in English for the reader’s convenience.

<sup>59</sup> Marti (1995) distinguishes between two different ideas in DRT: that the meaning (or “semantic value”) of a name is its referent, and the idea of Russellian singular propositions (in which the referent itself is a constituent of the proposition expressed). I shall focus here only on the former, less technical idea.

to assume that the only alternative to descriptivism is DRT – that these are the only possible options.)

Kaplan (1989 a, b) has been an important background figure in this development: he defined the notion of “directly referential,” but it is not entirely clear to what view exactly he wanted to commit himself, at least as far as proper names are concerned.”<sup>60</sup> Be that as it may, the full-blown DRT emerged with Almog (1984 and 1985), Salmon (1986), Wettstein (1986), and Soames (1987), for example (see also, e.g., Braun 1993, 1998, and 2001; Soames 2001).

In contrast, of the central figures of NTR, neither Kripke nor Putnam<sup>61</sup> or Devitt have subscribed to DRT. Kripke does later say that his own view is *closer* in various respects to Mill’s view than to the descriptivist tradition (see Kripke 1973/2013, 11; Kripke 1979, 239) – even that “a Millian line should be maintained *as far as is feasible*” (1979, 248; my emphasis). And Kripke does endorse the substitutivity of co-referential names in the contexts of *alethic* modalities.<sup>62</sup> However, he does *not* advocate universal substitutivity of names (e.g. in belief contexts), *nor* does he regard “Hesperus is Phosphorus” and “Hesperus is Hesperus” as interchangeable (see Kripke 1980, 20).<sup>63</sup>

Devitt has been a particularly important force at this issue. He has argued (Devitt 1989) that the revival of DRT is based on a problematic background assumption, a false dichotomy (he calls it “semantic presupposition”; abbreviated SP):

(SP) The *meaning* of a name is either *descriptive* or else it is the name’s *referent*.

Many seem to think that SP is true by definition or trivial. Devitt, however, questions this assumption. As a third alternative – he later called it “a shocking idea” – Devitt (1974, 1981, 1989, 2001, 2015) proposed that referring expressions do have, over and above the object denoted, a sort of sense – a way the referent is presented – but that this sense is not descriptive. Rather, he suggests that the causal-historical chain relevant for the name itself, which is often opaque to the particular language user, can play (at least in many respects) the role of the sense, or the meaning. Different types of causal-historical chains underlie “Hesperus” and “Phosphorus”, for example, and this explains their different roles in

---

<sup>60</sup> I think much the same can be said about Salmon’s early work (1981), in which he states, e.g., “it is neither helpful nor illuminating to see the central issue [with DRT] as a question whether proper names have sense” (ibid., 11).

<sup>61</sup> In Putnam’s case, I content myself with referring to Putnam (2001).

<sup>62</sup> That is, in sentence contexts involving metaphysical necessity or possibility.

<sup>63</sup> Kripke, however, raises some doubts as to whether the real historical Mill endorsed the latter, either.

reasoning and communication, and justifies the conclusion that they have distinct meanings.

What is the motivation for ascribing *meanings* to some strings of symbols (or of sounds), in any case? One plausible reason is to serve as an explanation of human behavior. Assume, for example, that Jason is told that Bob Dylan is in the room. This may trigger certain behaviors in him: for example, he might try to find Dylan and shake his hand. If Jason is told instead that Robert Zimmerman is in the room (and Jason does not know that Robert Zimmerman is in fact Bob Dylan), he may well react very differently. This supports the hypothesis that there is a difference of *meaning* in “Bob Dylan” and “Robert Zimmerman.”

One might now object that this suggestion would make meaning, or sense, much too fine-grained: does it not entail that any pair of different tokens of the same name, not to mention distinct but intuitively synonymous expressions, would always have different meanings? Not necessarily. First, Devitt’s proposal concerns types and not tokens.<sup>64</sup> Second, Devitt grants that similarity between causal-historical chains is a matter of degree (1981, 154–5).

I contend, for my part, that in practice, it is, to some extent, a matter of conventional stipulation how similar we require the chains to be for two expressions to be considered synonymous. The question is a bit tricky, especially with proper names: for example, “Germany” and “Deutschland” or “Finland” and “Suomi” are now commonly considered as synonymous pairs, although they probably have distinct origins. The idea of multiple grounding can perhaps be used to partly explain such cases. In other cases, like that of Saint Peter, the historical chains of Aramaic “Kepa” and Greek “Petros” (both meaning “rock”) presumably overlap, whereas his original name “Shimon” has a more distinct chain. Be that as it may, however, if the (multiple) groundings and the subsequent causal-historical chains of two co-referential expressions are largely distinct, this may be a sufficient reason to consider two expressions as non-synonymous.

Given how natural the idea of looking at the causal-historical chains is (once we have recognized them, at any rate), I find it puzzling how little explicit attention Devitt’s suggestion has received in the literature. Most philosophers have not given serious consideration to this alternative. Many do not seem to even be aware of the proposal. From the DRT camp, Salmon notes it but quickly dismisses it as “ill conceived if not downright desperate....wildly bizarre.... a confusion, on the order of a category mistake” (1986, 70–71).

---

<sup>64</sup> “Note that my view is not the genuinely preposterous view that the meaning of a name is a particular *token* causal link – and so is not open to Salmon’s ‘argument from subjectivity’” (Devitt 2012, fn 14).



However, allow me make an interesting historical observation: apparently very few have noted that in the original 1972 article version of “Naming and Necessity”, Kripke himself, though he perhaps did not unambiguously endorse the idea, did at least briefly mention it:

Hartry Field has proposed that, for some of the purposes of Frege’s theory, his notion of sense should be replaced by the chain which determines reference. (Kripke 1972, 346 n. 22)

This passage is, however, omitted from the reprinted 1980 book version.<sup>65</sup> As it happens, Kripke also mentioned the idea in passing in “A puzzle about belief”:

It has been suggested that the chain of communication ... might thereby itself be called a ‘sense’. Perhaps so (Kripke 1979, 248).

Furthermore, as Devitt (2015) notes, recently Kaplan, of all people – to many he is the father of contemporary DRT – has endorsed an idea very similar to this:

This might be an appropriate place to raise the question whether these arguments show that proper names are not Millian. If *Millian* means that different names of the same individual never differ semantically, I do not think that names are Millian, because I take the *way* the bearer is represented, even if nondescriptive, to belong to semantic theory. However, Mill himself claimed only that names had denotation but no connotation. Connotation was, for Mill, *descriptive* meaning that determines denotation. Mill believed that predicates and natural-kind terms had such connotations. So, if by *Millian* we mean that names do not have Millian connotations, then I do regard names as Millian since the *way* the bearer is represented is nondescriptive. As we have learned, it is important to separate *how* the individual is represented from the mechanism that determines *what* individual is represented. This is a distinction that the notion of a referential use of a definite description presupposes (Kaplan 2012, 167, endnote 22).

I believe that this idea of Devitt – whatever its limits – deserves broader interest. Though probably nothing can play all the roles that the traditional notion of meaning or Frege’s notion of sense was purported to play, it seems that if we accept the idea of the causal-historical chain of communication, it can play at least some of those roles. At any rate, Devitt’s proposal handles Frege’s puzzles quite nicely (see Devitt 1989). DRT, on the other hand, has made little indisputable progress with them.

## Acknowledgements

---

<sup>65</sup> At the 2013 Buenos Aires workshop (where both Devitt and I were present), Kripke explained that he had deleted the note simply because someone had informed him that he should have credited the idea to Devitt and not to Field.

Earlier versions (of parts) of this paper have been presented in London, Stirling, Florence, Hamburg, Helsinki and Turku. I would like to thank those who participated in the discussions on these occasions.

I have learned about these topics more from Michael Devitt than from anyone else — first from his writings, and later also from our innumerable discussions and from our correspondence. With this little piece, I want to congratulate my good friend Michael on the occasion of his eightieth birthday and, most of all, to celebrate his momentous life's work in philosophy.

## References

- Achinstein, P. (1968). *Concepts of Science. A Philosophical Analysis*. London: The Johns Hopkins University Press.
- Almog, J. (1984). Semantic anthropology. *Midwest Studies in Philosophy*, 9, 478–489.
- Almog, J. (1985). Form and content. *Noûs*, 19, 603–16.
- Almog, J., Perry, J., & Wettstein, H. (Eds.) (1989). *Themes from Kaplan*. New York: Oxford University Press.
- Bach, K. (1987). *Thought and Reference*. Oxford: Oxford University Press.
- Bach, K. (1998). Content: wide and narrow. In E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge.
- Braun, D. (1993). Empty names. *Noûs*, 27, 449–469.
- Braun, D. (1998). Understanding belief reports. *Philosophical Review*, 107, 555–595.
- Braun, D. (2001). Russellianism and explanation. In J. Tomberlin (Ed.), *Philosophical Perspectives*, 15. *Metaphysics* (pp. 253–89). Atascadero, CA: Ridgeview Publishing Company.
- Braun, D. (2006). Names and natural kind terms. In E. Lepore & B.C. Smith (Eds.), *The Oxford Handbook of Philosophy of Language* (pp. 490–515). Oxford: Oxford University Press.
- Bridges, J. (2010). Wittgenstein vs. contextualism. In A. M. Ahmed (Ed.), *Wittgenstein's Philosophical Investigations: A Critical Guide* (pp. 109–128). Cambridge: Cambridge University Press.
- Burge, T. (1979). Sinning against Frege. *Philosophical Review*, 88, 398–432.
- Burgess, J.P. (2013). *Kripke. Puzzles and Mysteries*. Key Contemporary Thinkers. Cambridge: Polity.
- Carnap, R. (1947). *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago: The University of Chicago Press.
- Chalmers, D. (2002). On sense and intension. *Philosophical Perspectives*, 16, 135–82.
- Chalmers, D. (2012). *Constructing the World*. Oxford: Oxford University Press.
- Church, A. (1950). On Carnap's analysis of statements of assertion and belief. *Analysis*, 10 (5), 97–99.
- Crane, T. (1991). All the difference in the world. *The Philosophical Quarterly*, 41, 1–25.

- Cumming, S. (2016). Names. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2016 Edition). URL = <<https://plato.stanford.edu/archives/fall2016/entries/names/>>.
- Currie, G. (1982). *Frege. An Introduction to His Philosophy*. Brighton: Harvester Press.
- Currie, G., & Eggenberger, P. (1983). Knowledge of meaning. *Noûs*, 17, 267–279.
- Davidson, D. (1990). The structure and content of truth. *Journal of Philosophy*, 87, 279–328.
- Devitt, M. (1974). Singular terms. *Journal of Philosophy*, 71, 183–205.
- Devitt, M. (1981). *Designation*. New York: Columbia University Press.
- Devitt, M. (1983). Dummett's anti-realism. *Journal of Philosophy*, 80, 73–99.
- Devitt, M. (1989). Against direct reference. *Midwest Studies in Philosophy*, 14, 206–240.
- Devitt, M. (1998). Reference. In E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge.
- Devitt, M. (2001). A shocking idea about meaning. *Revue Internationale de Philosophie*, v. 4, No. 218, 471–494.
- Devitt, M. (2006). Responses to the Rijeka papers. *Croatian Journal of Philosophy*, 6, 97–112.
- Devitt, M. (2008). Reference borrowing: a response to Dunja Jutronic. *Croatian Journal of Philosophy*, 8, 361–66.
- Devitt, M. (2012). Still against direct reference. In R. Schantz (Ed.), *Prospects for Meaning*, (pp. 61–84). Berlin: Walter de Gruyter.
- Devitt, M. (2015). Should proper names still seem so problematic? In A. Bianchi (Ed.), *On Reference* (pp. 108–143). Oxford: Oxford University Press.
- Devitt, M., & Sterelny, K. (1987). *Language and Reality*. Oxford: Basil Blackwell.
- Devitt, M., & Sterelny, K. (1999). *Language and Reality*. Second Edition. Oxford: Blackwell.
- Donnellan, K. (1970). Proper names and identifying descriptions. *Synthese*, 21, 335–358. Reprinted in D. Davidson & G. Harman (Eds.), *Semantics of Natural Language* (pp. 356–79). Dordrecht: Reidel, 1972.
- Dummett, M. (1973). *Frege: Philosophy of Language*. London: Duckworth.
- Dummett, M. (1974). The social character of meaning. Published in Dummett 1978, 420–430.
- Dummett, M. (1978). *Truth and Other Enigmas*. London: Duckworth.
- Dummett, M. (1993). *The Seas of Language*. Oxford: Oxford University Press.
- Dupre, J. (1981). Natural kinds and biological taxa. *Philosophical Review*, 90, 66–90.
- Evans, G. (1973). The causal theory of names. *Proceedings of the Aristotelian Society*, Suppl. vol. 47, 187–208.
- Evans, G. (1985). Understanding demonstratives. In G. Evans, *Collected Papers*, 291–321. Oxford: Clarendon Press.
- Everett, A. (2005). Recent defenses of descriptivism. *Mind & Language*, 20, 103–139.
- Fitch, G. W. (2004). *Saul Kripke*. Bucks: Acumen.
- Frege, G. (1892a). On sense and meaning. In McGuinness 1984, 157–177.
- Frege, G. (1892b). On concept and object. In McGuinness 1984, 182–194.
- Frege, G. (1914). Letter to Jourdain. In G. Frege, *Philosophical and Mathematical Correspondence* (ed. B. McGuinness) (pp. 78–80). Oxford: Blackwell.
- Heck, R., & May, R. (2006). Frege's contribution to philosophy of language. In E. Lepore & B.C. Smith (Eds.), *The Oxford Handbook of Philosophy of Language* (pp. 3–39). Oxford: Oxford

University Press.

- Jackson, F. (1998). Reference and description revisited. *Philosophical Perspectives*, 12, 201–218.
- Kaplan, D. (1989a). Demonstratives: an essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In Almog et al. 1988, 481–563.
- Kaplan, D. (1989b). Afterthoughts. In Almog et al. 1989, 565–614.
- Kaplan, D. (2012). An idea of Donnellan. In J. Almog & P. Leonardi (Eds.), *Having in Mind: The Philosophy of Keith Donnellan* (pp. 122–175). New York: Oxford University Press.
- Katz, J.J. (1990). Has the description theory of names been refuted? In G. Boolos (Ed.), *Meaning and Method: Essays in Honour of Hilary Putnam* (pp. 31–61). Cambridge: Cambridge University Press.
- Katz, J.J. (1994). Names without bearers. *Philosophical Review*, 103, 1–39.
- Kremer, M. (2010). Sense and reference: the origins and development of the distinction. In T. Ricketts & M. Potter (Eds.), *The Cambridge Companion to Frege* (pp. 220–292). Cambridge: Cambridge University Press.
- Kripke, S. (1971). Identity and necessity. In M.K. Munitz (Ed.), *Identity and Individuation* (pp. 135–164). New York: New York University Press.
- Kripke, S. (1972). Naming and necessity. In D. Davidson & G. Harman (Eds.), *Semantics of Natural Language* (pp. 253–355). Dordrecht: Reidel.
- Kripke, S. (1973/2013). *Reference and Existence: The John Locke Lectures*. Oxford: Oxford University Press.
- Kripke, S. (1979). A puzzle about belief. In A. Margalit (Ed.), *Meaning and Use* (pp. 239–83). Dordrecht: Reidel.
- Kripke, S. (1980). *Naming and Necessity*. Reprint of (Kripke 1972) with a new introduction. Cambridge, MA: Harvard University Press.
- Kripke, S. (2008). Frege's theory of sense and reference: some exegetical notes. *Theoria*, 74, 181–218.
- Kroon, F. (1987). Causal descriptivism. *Australasian Journal of Philosophy*, 65, 1–17.
- Lewis, D. (1984). Putnam's paradox. *Australasian Journal of Philosophy*, 62, 221–36.
- Lycan, W. (2006). Names. In M. Devitt & R. Hanley (Eds.), *The Blackwell Guide to Philosophy of Language* (pp. 255–273). Oxford: Blackwell.
- Lycan, W. (2008). *Philosophy of Language. A Contemporary Introduction*. 2<sup>nd</sup> Edition. Oxon: Routledge.
- Martí, G. (1995). The essence of genuine reference. *Journal of Philosophical Logic*, 24, 275–289.
- May, R. (2006). The invariance of sense. *Journal of Philosophy*, 102, 111–44.
- McDowell, J. (1977). On the sense and reference of a proper name. *Mind*, 86, 159–85.
- McGinn, C. (1982). The structure of content. In A. Woodfield (Ed.), *Thought and Content* (pp. 207–258). Oxford: Oxford University Press.
- McGuinness, B. (ed.), (1984). Gottlob Frege, *Collected Papers on Mathematics, Logic, and Philosophy*. Oxford: Basil Blackwell.
- Millar, A. (1977). Truth and understanding. *Mind*, 86, 405–416.
- Miller, A. (2006). Realism and anti-realism. In E. Lepore & B.C. Smith (Eds.), *Handbook of the Philosophy of Language* (pp. 983–1005). Oxford: Oxford University Press.

- Noonan, H.W. (2001). *Frege: A Critical Introduction*. Cambridge: Polity Press.
- Papineau, D. 1979). *Theory and Meaning*. Oxford: Clarendon Press.
- Plantinga, A. (1974). *The Nature of Necessity*. Oxford: Oxford University Press.
- Putnam, H. (1962). The analytic and the synthetic. In H. Feigl & G. Maxwell (Eds.), *Scientific Explanation, Space, and Time*. Minnesota Studies in the Philosophy of Science Vol. 3 (pp. 358–397). Minneapolis: University of Minnesota Press. Reprinted in Putnam 1975d, 33–69.
- Putnam, H. (1965). How not to talk about meaning. In R.S. Cohen & M.R. Wartofsky (Eds.), *Boston Studies in the Philosophy of Science, Vol. 2* (pp. 205–222). New York: Humanities Press. Reprinted in Putnam 1975d, 117–131.
- Putnam, H. (1970). Is semantics possible? Reprinted in Putnam 1975d, 139–52. (Originally published in Howard Kiefer & Milton Munitz (eds.), *Languages, Belief and Metaphysics*, Albany: SUNY Press, 1970.)
- Putnam, H. (1973). Explanation and reference. In G. Pearce & P. Maynard (Eds.), *Conceptual Change*. Dordrecht: Reidel. Reprinted in Putnam 1975d, 196–214.
- Putnam, H. (1974). Comment on Wilfrid Sellars. *Synthese*, 27, 445–455.
- Putnam, H. (1975a). The meaning of ‘meaning’. In K. Gunderson (Ed.), *Language, Mind, and Knowledge*, Minnesota Studies in the Philosophy of Science VII (pp. 131–193). Minneapolis: University of Minnesota Press. Reprinted in Putnam 1975d, 215–271.
- Putnam, H. (1975b). Language and reality. In Putnam 1975d, 272–290.
- Putnam, H. (1975c). The refutation of conventionalism. In M. Munitz (Ed.), *Semantics and Meaning*. New York: New York University Press. Reprinted in H. Putnam 1975d, 153–191.
- Putnam, H. (1975d). *Mind, Language and Reality: Philosophical Papers*, vol. 2. Cambridge: Cambridge University Press.
- Putnam, H. (1986). Meaning holism. In L.E. Hahn & P.A. Schilpp (Eds.), *The Philosophy of W. V. Quine* (pp. 405–431). La Salle, Ill.: Open Court.
- Putnam, H. (1988). *Representation and Reality*. Cambridge, MA: MIT Press.
- Putnam H. (2001). Reply to Michael Devitt. *Revue internationale de philosophie* 4/2001 (n° 218), 495–502.
- Quine, W.V. (1960). Carnap and logical truth. *Synthese*, 12, 350–374.
- Quine, W.V. (1963). Necessary truth. Reprinted in W.V. Quine, *The Ways of Paradox and Other Essays*. Revised and enlarged edition (pp. 68–76). Cambridge: Harvard University Press, 1976.
- Raatikainen, P. (2005). On how to avoid the indeterminacy of translation? *The Southern Journal of Philosophy*, 43, 395–414.
- Raatikainen, P. (2006). Against causal descriptivism. *Mind and Society*, Vol. 5, No 1, 78–84.
- Raatikainen, P. (2010). The semantic realism/anti-realism dispute and knowledge of meanings. *The Baltic International Yearbook of Cognition, Logic and Communication*, Vol. 5, 1–13.
- Reimer, M. (2009). Reference. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2009 Edition).  
URL = <<http://plato.stanford.edu/archives/sum2009/entries/reference/>>.
- Russell, B. (1918). The philosophy of logical atomism. In R.C. Marsh (Ed.), *Logic and Knowledge* (pp. 177–281). London: Allen & Unwin, 1956.

- Salmon, N. (1981). *Reference and Essence*. Princeton: Princeton University Press.
- Salmon, N. (1986). *Frege's Puzzle*. Cambridge, MA: MIT Press.
- Schwartz, S.P. (1977). Introduction. In S.P. Schwartz (Ed.), *Naming, Necessity, and Natural Kinds* (pp. 13–41). Ithaca & London: Cornell University Press.
- Searle, J. (1958). Proper names. *Mind*, 67, 166–173.
- Searle, J. (1967). Proper names and descriptions. In P. Edwards (Ed.), *Encyclopedia of Philosophy*, Vol. 6 (pp. 487–491). New York: MacMillan.
- Searle, J. (1971). Introduction. In J. Searle (Ed.), *The Philosophy of Language* (pp. 1–12). Oxford: Oxford University Press.
- Searle, J. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- Segal, G. (2000). *A Slim Book About Narrow Content*. Cambridge, MA: MIT Press.
- Shapere, D. (1964). The structure of scientific revolutions. *The Philosophical Review*, 73, 383–94.
- Shapere, D. (1966). Meaning and scientific change. In R. Colodny (Ed.), *Mind and Cosmos: Essays in Contemporary Science and Philosophy* (pp. 41–85). Pittsburgh: University of Pittsburgh Press.
- Soames, S. (1987). Direct reference, propositional attitudes, and semantic content. *Philosophical Topics*, 15, 47–87.
- Soames, S. (2001). *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. Oxford and New York: Oxford University Press.
- Stanford, P.K., & Kitcher, P. (2000). Refining the causal theory of reference for natural kind terms. *Philosophical Studies*, 97, 97–127.
- Sterelny, K. (1983). Natural kind terms. *Pacific Philosophical Quarterly*, 64, 110–25.
- Strawson, P.F. (1959). *Individuals*. London: Routledge.
- Strawson, P.F. (1969). Meaning and truth. Reprinted in P.F. Strawson, *Logico-Linguistic Papers*, 1971, (170–89). London: Methuen.
- Travis, C. (1989). *The Uses of Sense: Wittgenstein's Philosophy of Language*. Oxford: Clarendon Press.
- Unger, P. (1983). The causal theory of reference. *Philosophical Studies*, 43, 1–45.
- Wettstein, H. (1986). Has semantics rested on a mistake? *Journal of Philosophy*, 83, 185–209.