Temporal Externalist Descriptivism on Natural Kind Terms: Beyond the Causal–Historical Analysis

Haruka Iikawa (Hitotsubashi University)

Go Sasaki (Independent Researcher)

Abstract

The traditional debate over theories of reference of natural kind terms faces a serious dilemma. On the one hand, although direct reference theory, or the causal–historical analysis of reference to natural kinds, is still highly influential in the philosophy of language, there is a notorious "qua" problem: direct reference theory cannot uniquely determine the referents of natural kind terms. On the other hand, the standard descriptivism does not accommodate our externalist intuition. We propose temporal externalist descriptivism, where relevant future theorists determine references by past speakers. This position makes the externalist intuition about reference compatible with descriptivism about natural kind terms by ensuring the possibility of mistakes in the past and present practice. It also provides a defense of the controversial claim of temporal externalism based on a descriptivist analysis of reference. While causal–historical analysis focuses on inheritance through history, a reference has both a past and a future. In this sense, reference to natural kinds is essentially a provisional phenomenon.

1. Introduction

How are the meanings of natural kind terms determined? Entities known as natural kinds, such as water, oxygen, tigers, etc., are considered real because they exist in the world. Therefore, it is possible for individuals or even entire communities to be ignorant of or wrong about them. How, then, can we refer to natural kinds?

In the philosophy of language, direct reference theory (DR for short) has wielded significant influence since its proposal in the 1970s as an alternative to descriptivism regarding reference (Kripke, 1980; Putnam, 1975). While standard descriptivism (SD) claims that speakers associate a collection of descriptions with a term, and this determines the reference, DR asserts that the reference is determined without mediation by such descriptions.[1] DR's most compelling and influential form is the causal theory of reference or so-called causal-historical analysis. According to this account, once a reference has been fixed in a "naming ceremony," speakers' descriptions play no constitutive role in reference. This is considered DR's primary advantage: it can eliminate the myth of reference described above. If DR is accurate, we can refer to natural kinds despite our ignorance because speakers' descriptions are not relevant to reference.

Despite this advantage and its enduring popularity in the philosophy of language, DR fails to explain how successful reference by ignorant speakers is possible due to the so-called *qua* problem (see Crane, 2021; cf. Devitt & Sterelny, 1999; Raatikainen, 2021). Since DR does not appeal to full-fledged descriptions of natural kinds, it cannot uniquely determine references. It cannot explain why our use of terms refers to a natural kind, not to other groups that are similar but not natural. This indeterminacy motivates some form of descriptivism.

We argue that temporal externalism (TE) can offer such descriptivism. TE, which has been defended using thought experiments, posits that events can partly determine the reference of

---

[1] Notice that these formulations of DR and descriptivism show that, in this paper, they are metasemantic theories, not semantic ones. Descriptivism in this sense is compatible with referential semantics which assigns referents as semantic values to names.

terms used before they occur (e.g., Jackman, 1999, 2020; Tanesini, 2014; Haukioja, 2020; Collins, 2006). Our temporal externalist descriptivism (TED) makes sense of successful reference by ignorant people. If our proposal is successful, it provides not only the best option for a metasemantics of natural kind terms but also a new argument for the controversial concept of TE.

This paper is structured as follows. Section 2 scrutinizes the pros and cons of DR. Its primary advantage is considered its ability to accommodate the externalist intuition that we can refer to something without knowing its essence. However, considering the *qua* problem, it fails to capture this intuition. Section 3 proposes TED, and Section 4 discusses three potential objections to it. These objections pertain to the reality of retrospective determination, the explanatory relevance of future events to reference, and the reality of natural kinds.[2]

2. The Externalist Intuition and the Deadlock in the Very Idea of Direct Reference to Natural Kinds

In this section, we roughly sketch DR and the *qua* problem. DR has two main response strategies in the literature, which fail to make sense of both the externalist intuition and the very notion of direct reference to natural kinds.

DR's main advantage over SD is that it appears capable of accommodating *the externalist intuition*. Putnam's famous Twin Earth thought experiment clearly illustrates this point (Putnam, 1975). Let us assume that there exists a planet that is an exact duplicate of Earth

---

[2] In this paper, we do not discuss the nature or definition of natural kinds. Instead, our focus is on understanding how the names of purportedly natural kinds *refer* to those kinds (such as to water, oxygen, tigers, etc.), regardless of how their naturalness is defined. We also do not explore the question of whether these supposedly natural kinds are truly natural or real in some sense. In Section 4, we briefly discuss how our theory of reference relates to the realism versus anti-realism debate and to the HPC theory of natural kinds in particular.

except for one aspect: on Twin Earth, there is twater, which is superficially indistinguishable from water but composed of XYZ instead of H_2O. Both earthlings and twin-earthlings were ignorant of the components of each substance they call "water," and they could not distinguish water from twater in the 16th century. However, semantic externalists argue, our intuition tells us that earthlings refer to water, and twin-earthlings to twater, by the same word: "water." DR seems to be able to accommodate this intuition while SD cannot. Since DR holds that ostension can fix references, those references can depend on the speakers' environment at the time of the naming ceremonies. While earthlings fixed and inherited a reference to the substance composed of H_2O, twin-earthlings fixed and inherited a reference to the substance composed of XYZ. SD faces a challenge in this scenario because, according to its principles, earthlings and twin-earthlings in the 16th century are supposed to associate the exact same descriptions with the term "water," yet they appear to be referring to two different substances. SD lacks the means to capture this referential difference. In other words, while DR seems immune to problems caused by speakers who are ignorant of or misunderstand referents, these cases pose significant challenges for SD.

However, DR also has serious problems with the very notion of "direct reference" to natural kinds. Indeed, even DR acknowledges that the first stage of fixing references plays a constitutive role. After all, the referent of a term depends entirely on what the first speaker(s) assigned to it in the naming ceremony. Nevertheless, the superficial properties relied upon in naming ceremonies are not sufficient to determine natural kinds. For example, if the referent of the term "gold" is identified with the qualities of glittering, it is indeterminate among various glittering metals, like platinum. This is the so-called *qua* problem (Devitt & Sterelny, 1999; Crane, 2021). To account for natural kind terms, DR must make sense of direct reference to referents *qua* these very natural kinds, such as gold, platinum, or water.

Notice that this problem arises even if other options of DR than Kripke's original causal-historical analysis are adopted. For example, Putnam (1975) argues that expert research is more

important than naming ceremonies as the starting point of a causal chain (see also Moreno, 2021). Although this position explains the fact that contemporary speakers can refer to H_2O rather than XYZ, it does not explain the referential success of the 16th-century speakers because even experts could not distinguish water from twater in that era. After all, the question is how ignorant speakers can fix their references to natural kinds, and every version of DR has to answer this question.[3]

There are two major responses to the qua problem. The first asserts that natural kinds have essences in some sense and these essences uniquely determine their references (*essentialism*). If the essence of each instance in front of each speaker assigns the reference by virtue of its existence in the world, speakers in naming ceremonies can successfully fix references even if they do not know, when aiming to refer to kinds, what their essences are.

This response, to be successful, requires a substantive metaphysical commitment that all individuals belong to a unique hierarchy of natural kinds (Crane, 2021). Most individuals belong to more than one kind: a tiger is a tiger, a mammal, an animal, etc. According to essentialism, then, they have multiple essences, each corresponding to a kind. Naming a kind seems to be indeterminate between these essences, and this indeterminacy requires the additional commitment to a unique hierarchy of natural kinds. If there is one, and if the first speakers in naming ceremonies intend a specific level of the hierarchy, their naming acts successfully determine unique natural kinds. For example, the natural kind term "tiger" is intended to refer to the level between Panthera and the Sumatran tiger. Even if the initial descriptions that speakers in a naming ceremony associate with a putative natural kind name

---

[3] Another example is Evans (1973), but even in this position the same problem would arise because it appeals to common knowledge shared by the community of speakers. (In the first place, however, it is a theory about proper names and not about natural kind terms, so this criticism may be unfair to that position.)

are somewhat flawed, the term can refer to the kind if the speakers specify the level of the relevant hierarchy at which the kind is located.

However, this conception of *kind* does not conform so well to scientific practices (Crane, 2021). In actual sciences, kinds are considered relative to theories or disciplines. In some cases, scientists even refer to a cross-hierarchical kind. For example, Newtonian fluid is a cross-hierarchical kind because it belongs to the hierarchies of both physics and chemistry. If natural kinds constitute any unique hierarchy, such cross-hierarchical kinds are not natural kinds. This suggests that such a hierarchical conception of natural kinds diverges from scientific kinds.

Moreover, essentialism removes the main advantage of DR. Kind terms do not necessarily track any essences of individuals, and they may fail to refer to natural kinds. For example, since there is great diversity among fishes and the kind term "fish" may not capture any essence they have, it may not refer to a natural kind. "Phlogiston" is an obvious example of a kind term that does not track any essences of substances. DR with essentialism suggests that references to natural kinds are somehow automatically determined on the side of the world, meaning that speakers' descriptions are not required to track the essences of kinds. Indeed, earthlings in the 16th century did not know the essence of water, if that essence is that water is mainly composed of H_2O.

The cases of "fish" and "phlogiston" clearly demonstrate that there is no guarantee that the act of naming a kind, performed by an ignorant speaker, will succeed in capturing an essence and referring to a natural kind. So, what determines whether an act of naming can successfully capture an essence and refer to a natural kind? As Raatikainen (2021) rightly points out, DR can appeal to the first speakers' descriptions or intentions. However, the Twin Earth thought experiment demonstrates that these descriptions or intentions are insufficient to determine whether the newly invented term can capture an essence and refer to a natural kind. Here, DR with essentialism seems to face a serious dilemma: on one horn, if DR does not require the first speakers in a naming ceremony to track the essence of the putative referent of

a natural kind, it does not seem to be able to explain when reference to a natural kind is successful. On the other horn, if DR *does* require the tracking condition, it fails to accommodate the externalist intuition.[4]

Let us now turn to the second and seemingly more promising response to the *qua* problem, which suggests the possibility of "partial reference." Even if the reference of a kind term is not fully determined, the term can partially refer to a kind insofar as the superficial descriptions that were associated with the term in the naming ceremony, and that fix its reference, do track the kind (Raatikainen, 2021; see also Field, 1973; Devitt, 1981; M. Wilson, 1982). According to Field (1973), whether Newton's use of "mass" refers to relativistic mass or proper mass cannot be determined. Field persuasively argues that this indeterminacy does not prevent a Newtonian use of "mass" from successfully referring: it partially refers to both relativistic mass and proper mass. If this line of response is sound, DR could encompass a variety of concepts of the kinds used in scientific practice (Raatikainen, 2021).

Wilson (1982) also argues that the reference of a term is partially indeterminate because it depends on *the range of applicability.* Just as there are conditions under which thermometers function properly, there are also ranges of application of words. For Wilson's example, the "druids," who drifted to a desert island in ancient times, speak a language similar to old English, but for the first time they witness an airplane and start calling it a "bird." This application of "bird" is hardly expected before because if that plane had been found in a crashed condition, they would have called it a "house," not a "bird." Wilson argues that this kind of case is explained as the expansion or evolution of the range of applicability of a term. The extension assignment of the word "bird" in the druids' language to Aves, before their encounter with the airplane, should be relative to a range of objects similar to those they have encountered.

---

[4] Crane (2021) distinguishes the variants of essentialism more finely and examines each critically.

Whether the word "bird" in their language is applicable to objects beyond that range, such as airplanes, is indeterminate at that time. By the same token, DR theorists might say that it is indeterminate whether the referent of "water" in 16th-century language is water or twater, and our language has evolved to refer to just water. Indeed, it is possible, and perhaps even useful to identify semantic facts in this way.

However, these responses are again insufficient to save DR's greatest advantage: being able to explain the externalist intuition. Since, in the 16th century, earthlings did not have knowledge about the chemical composition of water, they did not distinguish water from twater. Thus, if they only partially referred to water, their kind term "water" partially referred to *both* water and twater, just as Newton's use of "mass" partially referred to both relativistic and proper mass. This amounts to giving up the externalist intuition that earthlings in the 16th century referred to water, not twater. Even Wilson's evolution view, after all, cannot explain the "this is water" statement about XYZ in the 16th century or the "this is gold" statement about platinum in that era as *false* rather than *indeterminate*. In this view, we could not assess the truth of these statements even if we know that water is mainly composed of H_2O and gold consists of Au.

In sum, the biggest putative advantage of DR (and the biggest challenge for SD) is the ability to explain the externalist intuition; on the other hand, the strongest argument against DR is the *qua* problem, which implies that DR does not make sense of the very idea of direct reference to natural kinds. DR's two major lines of response to the *qua* problem do not definitively establish the concept of direct reference to natural kinds, nor do they fully preserve DR's greatest advantage—its ability to accommodate the externalist intuition.

3. Temporal Externalist Descriptivism

The lesson from the *qua* problem is that in referencing natural kinds, some descriptive elements are required. These descriptions should not be merely formal: so-called "causal descriptions,"

indicating causal relationships between kinds and terms, are insufficient. Ultimately, this formal type of description faces the same *qua* problem as DR does. We need more substantive descriptivism, which both overcomes the *qua* problem and accommodates the externalist intuition.

Psillos proposes a descriptivist theory that could serve as a prototype. This theory is composed of two theses (Psillos, 1999, Chapter 12). (1) The success condition: a term *t* refers to an entity *x* if and only if *x* satisfies the core causal description associated with *t*. (2) The identity condition: two terms *t* and *t'* denote the same entity if and only if the core causal description of *t'* takes up the kind-constitutive properties of the core causal description associated with *t*.[5]

These core causal descriptions are descriptions that denote the kind-constitutive properties that purportedly explain or ground various superficial properties of the kinds.[6] For example, a property of being (mainly composed of) H_2O explains various other properties of water, such as being liquid at room temperature, being almost colorless and transparent, producing

---

[5] Psillos requires an additional condition: that their putative referents play the same causal role with respect to a network of phenomena. However, we omit this because it seems a simple consequence of condition (2).

[6] Psillos (1999, p. 277) states the following about kind-constitutive properties:

Generalizing on Enç's ideas, one may say that a kind-term refers to a natural kind by virtue of the fact that the body of information which is typically associated with a kind-term has its causal origin in the kind-constitutive properties of the kind. This means that this information has the propositional content it does because the kind has the kind-constitutive properties it does. For instance, to say that the causal origin of the body of information associated with the term "water" lies in the chemical constitution of water is to say the following: if the liquid which is colourless, odourless, still, thirst-quenching, etc., and whose boiling point is 100° Celsius, its freezing point 0° Celsius etc., were not H_2O it would not have these manifest properties, and the propositional content of the information associated with the term "water" would be different.

hydrogen and oxygen when an electric current is applied, and so on. Notice that core causal descriptions are different from formal causal descriptions. Merely formal causal descriptions, indicating that the causal sources of the uses of names are such and such, cannot uniquely determine kinds as their referents insofar as causal–historical analysis cannot (for a defense of such formal descriptivism, see Poller, 2020). In contrast, core causal descriptions are substantive in the sense that they are expected to unify various descriptions of superficial properties into a unified portrayal of a single kind. The candidates for these core causal descriptions include descriptions of chemical and physical components, genetic and phylogenetic descriptions, and so on.

The two conditions of Psillos' descriptivism fit well into scientific practice in two respects (Psillos, 1999, Chapter 12). First, the success condition can explain the failure of empty names to refer. "Phlogiston" refers to nothing because there is no substance satisfying the core causal descriptions associated with it, such as those stating that phlogiston causes combustion and is emitted during combustion.

Second, the identity condition could, at least partially, explain continuity of reference. According to Psillos, the old-fashioned technical term "luminiferous ether," used in 19th-century physics, refers to what the "electromagnetic field" refers to in contemporary physics. "Luminiferous ether" is associated with, according to Psillos, the core causal description that it is the medium through which light propagates, and this description is shared with "electromagnetic field."[7] Thus, the identity condition is satisfied.

These examples suggest that Psillos' descriptivism offers a more or less plausible position regarding the references of natural kind terms. It succeeds, to some extent, in explaining the continuity of these references by combining descriptivism with causal elements. The position

---

[7] An anonymous reviewer notes that the core status of this description is debatable. However, we argue that our improvements to Psillos' theory resolve this arbitrariness about the core status because, in our position, later scientists decide what is the core. See also footnote 12.

also encompasses the characteristics of scientific kinds that depend on their background disciplines or theories.

Despite these improvements, Psillos' descriptivism faces the same problem as SD. This is because some of the core causal descriptions that people have formerly associated with natural kind terms are false. For example, although Lavoisier is well known as one of the discoverers of oxygen, he associated with "oxygen" the core causal description that it is the element or source of all acids. This description is false, but we still think of him as one of the discoverers of oxygen. This thought implies that Lavoisier's use of "oxygen" successfully referred to oxygen. Thus, even Psillos' version of descriptivism falls prey to the objection from the externalist intuition.

We can overcome this objection by incorporating *temporal externalism* (TE) into Psillos' descriptivism. TE claims that later events determine the meaning of earlier language use (for the original version of TE, see Jackman, 1999, cf. 2020; Tanesini, 2014; Haukioja, 2020; for an application to natural kind terms, see Collins, 2006; for a slightly different conception of TE, see Reinikainen, 2022). This concept seems odd at first glance. Many advocates of TE (e.g., Tanesini, 2014; Haukioja, 2020) use the following thought experiment. In the 18th century, gold was identified by its property of being insoluble in all but aqua regia. Since platinum shares this property, the description "insoluble in all but aqua regia" does not determine the referent of "gold." Platinum was discovered in Peru in the 18th century and was distinguished from gold. However, we could have chosen to use the term "gold" to refer to platinum as well as gold. If we had chosen this usage, the advocates of TE argue, any usage of the term "gold" *before* the 18th century would have referred to platinum as well as gold. This intuition, if sound, shows that a later decision about the usage of "gold" would then determine whether the earlier

use of "gold" had referred to platinum, and supports TE.[8] Instead of arguing for this thought experiment, we defend TE by demonstrating that it can support a satisfactory explanation of the reference of natural kind terms (see Collins, 2006).

TE can update Psillos' descriptivism as follows:

The success condition: a kind term *t* refers to an entity *x* if and only if *x* uniquely satisfies the core causal descriptions *the relevant theory in the future* associates with *t*.

The identity condition: Two kind terms *t* and *t′* denote the same entity if and only if (1) (a) past speakers of *t′* associate it with a set of descriptions *D* that a relevant future theory also associates with *t* and (b) the future theory associates with *t* the core causal descriptions which explaining *D* denotes, or (2) there is a third term t″ such that both of *t* and *t′* satisfy condition (1) for it.

The success condition explains the case of Lavoisier's "oxygen" well. Lavoisier's use of "oxygen" referred to oxygen successfully because the entity satisfies the core causal descriptions that contemporary chemistry associates with the term: for example, it is composed of O_2. The identity condition ensures that Lavoisier's use of "oxygen" and our own use of it denote the same entity. Lavoisier associated descriptions with the term our chemistry also endorses (e.g., the substance is produced by such and such an experiment), and the properties these descriptions denote are explained by the core causal descriptions of our chemistry.[9]

The question becomes: which later theories or descriptions are *relevant* to the meanings of earlier terms? How and why is present chemistry relevant to determining the reference of

---

[8] Noting that TE is based on the crucial assumption that the referents of earlier and later uses of relevant linguistic items are identical. We discuss this assumption in section 4.

[9] Similarly, we can say that Priestley's "dephlogiston air" also refers to oxygen and makes Priestley one of the discoverers of oxygen.

Lavoisier's "oxygen?" DR provides an important insight in this regard. It argues that social interactions between speakers ground the inheritance of their references. This suggests that the future and past uses of terms are relevant insofar as later speakers are disposed to think both their own use of each term and the previous uses refer to the same object. This in turn means that later theories or descriptions are relevant to earlier references if and only if the later speakers are disposed to implicitly take the identity condition to be satisfied.

This characterization of relevance fits well with Psillos' example of how Einstein reinterpreted "luminiferous ether." Einstein (according to Psillos) took the referent of the term to be identical to that of "electromagnetic field" because he thought the two terms played the same role in each background theory and that his own electromagnetic field theory could explain that role.

Further, the proposed version of *temporal externalist descriptivism* (TED) is immune to the objection from the externalist intuition.[10] Lavoisier indeed misunderstood oxygen and falsely described "oxygen" as being the source of all acids. Earthlings in the 16th century were indeed too ignorant of the essence of water to distinguish it from twater. However, as the success condition states, references by ignorant speakers in earlier periods are determined by the core causal descriptions that later science associates with terms if the identity condition is satisfied. The identity condition says, in turn, that Lavoisier's or the 16th-century earthlings' false descriptions do not affect the continuity of reference between their use of "oxygen" or "water" and our chemistry.

This is because continuity is guaranteed by *D*, the set of descriptions *commonly* held both by Lavoisier or by 16th-century earthlings and also by our chemistry. Since later theorists' attitudes secure the relevance of their theories to the references of terms' earlier uses, which

---

[10] Collins (2006) has already pointed out that TE for natural kind terms can meet the challenge of the externalist intuition. One of our unique contributions is to show that TE can reinforce descriptivism through the concrete formulation of TED.

Accepted Version

descriptions *D* includes depends on later speakers' intentions. Meanwhile, our contemporary chemistry is relevant to earthlings' past use of "water," but not to twin-earthlings' "twater," because our chemists are disposed to take water on Earth to be water but not to take twater on Twin Earth as the referent of their own use of "water." In this sense, which later theories are relevant or identical to those of the past is indeed contingent. However, this contingency is one of the very points of TE. Using various thought experiments, supporters of TE have tried to show that earlier meanings are sometimes contingent on relevant later speakers' decisions.[11] In a nutshell, two facts secure the continuity of reference: (1) there is a *relevant* (in the above sense) set *D* and (2) the core causal descriptions of our chemistry (e.g., oxygen consists of O_2, water is mainly composed of H_2O) explain the properties *D* denotes.[12] Once its continuity is ensured, the success of a reference is determined by relevant future theories. Therefore, cases of misunderstanding or ignorance do not threaten TED.

In addition, TED suggests an important lesson about the very phenomenon of reference: it is relative to its background theory. Since our present science contains some errors, it is always subject to revision, so reference to natural kinds is essentially a provisional phenomenon.

4. The Problem of Determination from the Future and the Challenge of Realism

---

[11] Our version of TED can also explain the thought experiment about gold and platinum. Although the description of "gold" people in the 18th century associated with the term does not determine the referent, modern chemists are disposed to take the term to be identical to their "gold," not "platinum." This contingent (disposition toward a) decision provides relevance, ensuring the continuity of reference of "gold" in different eras.

[12] Notice that in our view, it is up to later scientists to determine what is the core, i.e., which descriptions integrate a bunch of other descriptions, and descriptions shared with people in the past need not be core causal descriptions. This means that even if the shared descriptions are not the core, they can still be relevant.

Finally, let us respond to three possible criticisms and develop TED a bit further. The first two criticisms are directed against TE in the literature; the other is an assumed criticism drawn from realism about natural kinds.

Recently, Reinikainen (2022) critically examined the most representative literature on TE and offered two criticisms of the concept of retrospective determination.[13] As we understand them, the criticisms are the following: the idea of TE (1) fails to satisfy the Wittgensteinian criterion of adequacy and (2) fails to provide any explanatory surplus to language use in the past.

By "Wittgensteinian criterion," Reinikainen means:

> For one, it is one thing to claim that we often take the past community members to be governed by the same conceptual norms as we are and another to claim that the past community members *really* are governed by the norms as we enforce them now. Although Tanesini does not well articulate what exactly she means by "institution of norms", there is at least one necessary criterion of adequacy any such account should meet, which is that the norms may be incorrectly enforced. So, she owes an account explaining what it would be for our retroactive institution or enforcement of conceptual norms to *fail* to hold the past members accountable to what we take to be the true, or correct, meaning of the word.

In sum, the Wittgensteinian criterion is that there should be a distinction between the following two states of affairs: one in which later interpreters merely think that the meanings of earlier terms are determined in a specific manner, and another in which meanings are actually

---

[13] Although the paper is mainly about the relationship between Kripkenstein's rule-following paradox (Kripke, 1982) and TE, its criticisms of TE are independently intelligible.

determined in that manner. This distinction amounts to saving the possibility that later interpretations can fail to determine the references of earlier generations' terms. Reinikainen effectively criticizes various attempts to argue for TE by pointing out that they all fail to satisfy this criterion.

Nevertheless, our version of TED *can* satisfy this criterion because later speakers can fail to determine the references of earlier terms in at least two ways. (1) They can fail to satisfy the success condition, i.e., they can associate core causal descriptions with a term that are simply false, or insufficient to determine a unique kind. In such cases, the term fails to refer to any entity (until further future theorists associate a sufficiently determinate description with it and the success condition obtained). (2) The later speakers can fail to satisfy the identity condition, misidentifying the descriptions earlier language users associated with a term. For example, it is possible that Lavoisier did not perform the oxygen isolation experiment *e*, and that the experiment *e′* he actually performed was one in which oxygen could not be isolated. In this case, our present chemistry fails to determine the reference of Lavoisier's use of "oxygen" despite our interpreter's intention to make our chemistry relevant to Lavoisier's "oxygen."

Although Reinikainen (2022) further suggests that the distinction the Wittgensteinian criterion calls for should be real, in the sense of being entirely mind-independent, this is too strong a demand. The distinction between success and failure is one thing; mind-independence or reality is another. Our version of TED can secure the distinction in a partially mind-dependent way, but this is innocent *per se*. We will return to this point at the end of this section.

Let us now turn to Reinikainen's (2022) second criticism. What contribution can later theories make to explaining people's past linguistic behaviors? Reinikainen (2022) attempts to show, using some thought experiments, that one's mental state at a certain time and one's prior behaviors are together sufficient to explain one's previous linguistic behaviors. The idea is that later decisions and actions may *change* the meanings of linguistic items but do not determine

meanings in the past. So, according to Reinikainen, subsequent behaviors cannot be explanatorily relevant to the past.

However, subsequent theories can contribute to explaining the past use of (at least) natural kind terms by allowing for an evaluation of whether that use referred successfully. If our arguments so far are sound, then only after settling the problem of the continuity or identity of references can the success or failure of reference be evaluated. Whether "luminiferous ether" successfully referred to anything depends on what core causal descriptions determine its reference. If Einstein's reinterpretation of "ether" as an electromagnetic field is successful, so that the referents of "luminiferous ether" and "electromagnetic field" are identical, and can be secured by the association of core causal descriptions from the theory of relativity with both terms secures their identity,[14] the old-fashioned term "luminiferous ether" successfully refers to an electromagnetic field.

In this way, subsequent theories are essential when identifying what past users of natural kind terms were talking about. As Sections 2 and 3 discussed, both atemporal descriptivism and any direct reference theory that relies solely on historical language use face various difficulties. If one understands the phenomenon of reference as atemporal, or if one believes that it is determined solely by past events, one must accept the incredible consequence that most statements made in pre-scientific eras that were apparently about natural kinds are, strictly speaking, not about natural kinds. Since ignorant speakers cannot fully determine the referents of natural kind terms, events of their time and before cannot make their statements about natural kinds proper. However, our externalist intuition tells us that even earthlings in the 16th century could make true or false, rather than indeterminate, statements about water as a natural kind,

---

[14] Strictly speaking, this reinterpretation changes not only the core causal descriptions but also which superficial descriptions are relevant and therefore included in D. For example, in Lorentz's ether theory, ether is said to produce the wind of ether, but the theory of relativity does not require this.

despite their incapacity to distinguish water from twater. It depends on relevant later theories whether earlier statements, such as statements that water is essential to life or that there must be luminiferous ether for light to propagate, are *true of* natural kinds.

It is time to respond to the third and final supposed objection from realism. Some realists may complain that descriptivist theories rely on people's mental state when defining references, and this may make it impossible to maintain scientific realism about natural kinds.

However, the issues of success and continuity of references, which we have been trying to secure, have often appeared as part of an argument defending realism from so-called pessimistic induction. Pessimistic induction is the most representative objection to realism: that since, in the history of science, there are many cases of empirically successful theories that later turned out to be false, even currently successful theories may later turn out to be wrong. One of the main refutation strategies of the realists is as follows. Yes, past scientific theories contained errors, but the crucial aspects of scientific theories have been handed down from the past to the present (see Chakravartty, 2017, Section 3.3). To that extent, scientific theory has (at least partially) captured reality. This line of counterargument hinges on the assumption that there is continuity of references.[15] Our arguments for TED can reinforce the realist strategy by

---

[15] Note that on the one hand, in the philosophy of science, there is an independent tradition since Kuhn that questions the continuity of references or topics in scientific theories (see Kuhn, 2012). On the other hand, in the context of the debate over scientific realism, even anti-realists tend to recognize some kind of continuity (e.g., that of aims, some part of theories, methodologies, and so on) against the Kuhnian tradition (see Laudan, 1986). The possible battleline of realists and anti-realists here is whether there is such continuity that ensures (1) the continuity of reference and (2) the reality of referents. Our position argues for the positive answer to the first question but remains neutral about the second. However, the positive answer to the second seems to plausibly presuppose the positive answer to the first. (Although other strategies give up the reality of referents and argue for different types of continuity and reality, such as structures, entities that can be manipulated, and so on, these cannot be discussed here.)

cashing out this assumption (remembering that Psillos is one of the staunchest defenders of scientific realism).

The compatibility of TED with realism is also suggested by the fact that TED is easily compatible with the homeostatic properties cluster (HPC) theory of natural kinds, which is one of the most influential realist positions in the literature. HPC theory, roughly, has two components: (1) each natural kind has an HPC and (2) HPCs are grounded by underlying causal mechanisms (Boyd, 1999; see also Bird & Tobin, 2023, Section 1.2). These seem to correspond to the two parts of the identity condition in our formulation of TED: (a) and (b). A set of descriptions *D* that both earlier speakers and a relevant later theory endorse, which condition (a) requires, can partially represent an HPC. In addition, since core causal descriptions, which (b) requires, should explain *D*, they can denote the underlying mechanism of the HPC.[16]

After all, TED can be neutral about the realist thesis that natural kinds are mind-independent. Of course, TED commits us to the thesis that references to natural kinds are mind-dependent in that they depend on subsequent theories. However, this need not mean that natural kinds are themselves mind-dependent. If realists acknowledge the additional commitment that there exists a worldly natural classification of kinds that is independent of our classification practices, they can make the thesis of mind-independent natural kinds compatible with a thesis of the mind-dependent reference of natural kind terms, such as TED.

In recent years, anti-realism about natural kinds has become increasingly influential (e.g., Papale & Montminy, 2023). TED could benefit this position as well because it can (at least

---

[16] The nature of such mechanisms and whether they really exist is controversial (Bird & Tobin, 2023, Section 1.2; see also Slater, 2015; Onishi & Serpico, 2022). However, although TED *can* accommodate homeostatic mechanisms, it does not commit us to their existence. It does commit us to the conditional claim that there should be a description in the future which, whether or not it is about a homeostatic mechanism, unifies various other descriptions that are associated with a term, and uniquely determines that term's referent if the term successfully refers to a natural kind.

partially) explain the folk-realist or externalist intuition that natural kinds exist independently of humankind's practices without any commitment of realism. After all, TED allows for the possibility that ignorant speakers successfully refer to natural kinds even though they associate false descriptions with natural kind terms. This amounts to saying that what determines the referents of natural kind terms is partially independent of people's (contemporary) practice.

In sum, TED claims that the references of natural kind terms depend on the minds or practices of relevant future communities. However, this does not deny our folk-realist intuition that we can always be wrong about natural kinds. On the contrary, TED can be seen even as a natural option for realists, cashing out some of their ideas as well as those of anti-realism.

5. Concluding Remarks

Finally, we have a theory of reference that reconciles two observations that are both plausible but in tension: that reference to natural kinds (1) requires descriptions that fully determine what they are and (2) is somewhat external to our minds. Traditional descriptivism can accommodate the first point, but fails to make sense of the second because it has only contemporary elements and lacks a historical point of view. Conversely, direct reference theory fits the second well, at least partially, but not the first, because it has retrospective elements but lacks a future point of view. The meanings of natural kind terms are determined not only by past and contemporaneous phenomena but also by the future.

References

Bird, A., & Tobin, E. (2023). Natural Kinds. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Spring 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2023/entries/natural-kinds/

Boyd, R. (1999). Homeostasis, Species, and Higher Taxa. In R. A. Wilson (Ed.), *Species* (pp. 141–186). The MIT Press. https://doi.org/10.7551/mitpress/6396.003.0012

Chakravartty, A. (2017). Scientific Realism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/

Collins, J. M. (2006). Temporal Externalsm, Natural Kind Terms, and Scientifically Ignorant Communities. *Philosophical Papers*, *35*(1), 55–68. https://doi.org/10.1080/05568640609485172

Crane, J. K. (2021). Two approaches to natural kinds. *Synthese*, *199*(5), 12177–12198. https://doi.org/10.1007/s11229-021-03328-9

Devitt, M. (1981). *Designation*. Columbia University Press. https://doi.org/10.7312/devi90836

Devitt, M., & Sterelny, K. (1999). *Language and Reality: An Introduction to the Philosophy of Language* (K. Sterelny, Ed.). MIT Press.

Evans, G., & Altham, J. E. J. (1973). The Causal Theory of Names. *Proceedings of the Aristotelian Society, Supplementary Volumes*, *47*, 187–225.

Field, H. (1973). Theory Change and The Indterminacy of Reference. *The Journal of Philosophy*, *70*(14), 462. https://doi.org/10.2307/2025110

Haukioja, J. (2020). Semantic burden-shifting and temporal externalism. *Inquiry*, *63*(9–10), 919–929. https://doi.org/10.1080/0020174X.2020.1805704

Jackman, H. (1999). We Live Forwards But Understand Backwards: Linguistic Practices and Future Behavior. *Pacific Philosophical Quarterly*, *80*(2), 157–177. https://doi.org/10.1111/1468-0114.00078

Jackman, H. (2020). Temporal externalism, conceptual continuity, meaning, and use. *Inquiry*, *63*(9–10), 959–973. https://doi.org/10.1080/0020174X.2020.1805706

Kripke, S. A. (1980). *Naming and Necessity: Lectures Given to the Princeton University Philosophy Colloquium*. Cambridge, MA, USA: Harvard University Press.

Kripke, S. A. (1982). *Wittgenstein on Rules and Private Language: An Elementary Exposition*. Harvard University Press.

Kuhn, T. S. (2012). *The Structure of Scientific Revolutions: 50th Anniversary Edition* (I. Hacking, Ed.). University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/S/bo13179781.html

Laudan, L. (1986). *Science and Values: The Aims of Science and Their Role in Scientific Debate*.

Moreno, L. F. (2021). Natural Kind Terms. In P. Stalmaszczyk (Ed.), *The Cambridge Handbook of the Philosophy of Language* (1st ed., pp. 283–299). Cambridge University Press. https://doi.org/10.1017/9781108698283.016

Onishi, Y., & Serpico, D. (2022). Homeostatic Property Cluster Theory without Homeostatic Mechanisms: Two Recent Attempts and their Costs. *Journal for General Philosophy of Science*, *53*(1), 61–82. https://doi.org/10.1007/s10838-020-09527-1

Papale, F., & Montminy, D. (2023). Natural Kinds: The Expendables. *Canadian Journal of Philosophy*, 1–18. https://doi.org/10.1017/can.2023.30

Poller, O. (2020). Causal Descriptivism. In *The Routledge Handbook of Linguistic Reference*. Routledge.

Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*. New York: Routledge.

Putnam, H. (1975). The Meaning of "Meaning." *Minnesota Studies in the Philosophy of Science*, *7*, 131–193.

Raatikainen, P. (2021). Natural kind terms again. *European Journal for Philosophy of Science*, *11*(1), 19. https://doi.org/10.1007/s13194-020-00344-3

Reinikainen, J. (2022). Meaning in time: On temporal externalism and Kripkenstein's skeptical challenge. *Synthese*, *200*(4), 288. https://doi.org/10.1007/s11229-022-03756-1

Slater, M. H. (2015). Natural Kindness. *The British Journal for the Philosophy of Science*, *66*(2), 375–411. https://doi.org/10.1093/bjps/axt033

Tanesini, A. (2014). Temporal Externalism: A Taxonomy, an Articulation, and a Defence. *Journal of the Philosophy of History*, *8*(1), 1–19. https://doi.org/10.1163/18722636-12341263

Wilson, M. (1982). Predicate Meets Property. *The Philosophical Review*, *91*(4), 549–589. https://doi.org/10.2307/2184801