Refining the Causal Descriptivist Theory for Scientific Terms[1]

Sabrina Hao[2]

Abstract

The referential status of theoretical terms has been at the center of the tension between scientific realism and theory change. On the one hand, descriptivist theories cannot account for referential success across theory changes; on the other hand, the causal theory of reference is accused of making reference too easy and fails to account for referential failure. Building on the works of Nola (1980) and Psillos (1999; 2012), I develop a refined version of the hybrid causal descriptivist theory in this paper. I first show how an exactly true causal description that survives through theory change is possible if we include the relevant experimental conditions and contexts. I then argue that the meaning of a scientific term fixed through a causal description consists of the referent only, which allows us to account for the complex linguistic behavior in the scientific community. At last, I draw from Donnellan's (1966) distinction between attributive use and referential use of definite description to account for scientific terms that are introduced without being associated with any phenomenon.

1 Introduction

The referential status of theoretical terms has been at the center of the tension between scientific realism and theory change. On the one hand, descriptivist accounts face the Kuhnian challenge of theory change. If a scientific term is defined by a theoretical description, then when the theory is superseded, it often turns out that nothing in reality fits the description, and the term fails to refer to anything. Since our best current theories are expected to be superseded as well, we are not justified in believing that the scientific terms in our current theories refer to anything. On the other hand, the causal theory of reference is accused of making reference too easy. If 'phlogiston' is defined as 'whatever causes combustion,' then 'phlogiston' would have been referring to oxygen, and 'phlogiston exists' would be trivially true. Responding to this challenge, Nola (1980), Cummiskey (1992), and Psillos (1999; 2012) respectively propose different versions of the hybrid causal descriptivist account, suggesting that on the basis of the purely causal account, a description of the referent's causal role in relation to the relevant observable phenomena should also enter the reference-fixing mechanism. While this approach succeeds in accounting for the referential failure of cases like 'phlogiston,' concerns remain about whether it can recover referential continuity in the history of science and account for the complex linguistic behaviors in scientific discourse.

---

[1] This paper is forthcoming in *Synthese*.

[2] Department of Philosophy, University of Pittsburgh, USA.

In this paper, I examine some major challenges to the causal descriptivist approach and develop a refined version of the causal descriptivist theory. The most prominent objection is that if the causal description involved in reference fixing of a scientific term is required to be exactly true, it faces the same problem as descriptivism, namely that nothing in reality could fit the description, and the term fails to refer. I will show how an exactly true causal description is possible if it includes detailed experimental contexts and conditions when the reference for a scientific term is fixed. Following that, I consider other challenges to the causal descriptivist account and provide a response to each. I show how the causal descriptivist approach, according to my proposal, can account for scientists' complex linguistic behaviors and establish referential continuity at least in commonly agreed cases.

The specific version of scientific realism I'm considering in this paper includes the claim that it is possible for us to gain ontological knowledge about unobservable entities and properties from current scientific success. A full defense of this claim requires a sophisticated argument against Laudan's (1981) famous objection that the referential success of central theoretical terms is not a necessary condition for a theory to be empirically successful (Psillos 1999; Ladyman 2011). This paper can by no means serve this purpose. The aim of this paper is to supply scientific realism with an adequate theory of reference, accounting for referential continuity such as 'oxygen' and 'electron,' as well as referential failure such as 'phlogiston.' Exactly how this semantic theory helps defend a version of scientific realism against Laudan's challenge would require future work.

With the underlying tie to scientific realism, the semantic account discussed in this paper only considers theoretical terms that are intended to refer to unobservable, non-ostendable entities in our scientific theories. While the discussion here says nothing about what a general semantic theory for natural kinds should be, it is also largely quarantined from the problems of classifying and referring to natural kinds in other domains of science (Stanford and Kitcher 2000; Okasha 2002; Slater 2015).[3]

The paper is structured as follows. In Section 2, I examine Putnam's original account for unobservable kinds and discuss whether a distinction between approximately true descriptions and completely false descriptions is viable. In Section 3, I show how an exactly true description that survives through theory change can be obtained by specifying experimental contexts. In Section 4, I argue that the meaning of a term fixed through causal description consists of the referent only and show how this can help account for scientists' complex linguistic behavior. In Section 5, I consider scientific terms that are introduced without being associated with any phenomenon, the reference of which thus cannot be fixed. I use Donnellan's (1966) discussion about the attributive use of definite description to demystify the puzzle.

2 Theory Change, Referential Failure, and Putnam's Account

[3] Recent descriptivist proposals against essentialism, such as Slater (2015) and Chakravartty (2007), recognize that most unobservable scientific entities are nevertheless essence kinds. Recent defenses of the purely causal theory, such as Deutsche (2023) and Reinikainen (2024), also recognize that theoretical terms need to be treated separately.

Descriptivist theories for scientific terms are vulnerable to the Kuhnian challenge regarding theory change (Kuhn 1970; Sanky 1994; Bird 2000). During theory changes, our theoretical description associated with a scientific term often changes a lot. In light of subsequent scientific development, it often turns out that nothing in reality fits the theoretical description in the superseded theories. If the theoretical description in the superseded theory fails to pick out anything in reality, we can hardly say that the successive theory is a better theory about the same thing (Putnam 1975). Failing to establish a convergence of ontological knowledge in scientific development raises doubt about any ontological commitment to the scientific terms in our best current theories, for we expect them to be superseded as well. This is what Putnam (1978, p.25) calls the argument of meta-induction: "Just as no term used in the science of more than fifty years (or whatever) ago referred, so it will turn out that no term used now (except maybe observation terms, if there are such) refers."

Scientific realists thus turn to the causal theory of reference, originally proposed by Putnam (1975; 1978) and Kripke (1980) and further developed by Devitt (1981; 1996). The causal theory of reference, in general, consists of two parts: reference fixing and reference borrowing. In the cases of observable kinds, reference fixing often occurs when a speaker comes into perceptual contact with a sample and introduces a term: "Let's call this sort of thing *x*". The term is thus fixed to refer to the kind of things that share the same internal structure as the sample. This event allows the reference to be fixed to designate a kind of thing, even though the description in the initial introduction act could be wrong. For instance, a speaker might introduce the term 'gold' in the following way: "Let's call this sort of yellow metal 'gold'" when they have a piece of gold in front of them. According to the causal theorists, it might very well be the case that gold turns out to be green, but the early usages of the term still succeed in referring to gold.

Reference borrowing occurs when speakers who did not participate in the reference-fixing events borrow the reference from speakers who did and further pass on the reference to other speakers. The underlying causal network established by each reference borrowing links each subsequent usage of the term back to the reference-fixing event, which links usages of the term to the sort of things being perceived.

For unobservable kinds, the causal theory of reference is often accused of making reference too easy that the reference of scientific terms cannot fail (Enć 1976; Nola 1980; Cummiskey 1992; Psillos 1999). According to the purely causal account, a reference to an unobservable kind is fixed in the way 'Let's call whatever sort of things that cause these effects *x*.' As Nola (1980, p. 509) points out, "theory seems to play no role" in the reference-fixing mechanism. 'Phlogiston,' for instance, is commonly agreed to be a case where the theoretical term fails to refer to anything in reality. However, if 'phlogiston' was introduced to refer to 'whatever sort of things cause combustion,' 'phlogiston' would successfully refer to oxygen. If scientific terms cannot fail to refer, scientific realists' ontological commitment to the referents becomes trivial. The success of our theories plays no role in such commitment.

However, I think this is a misinterpretation of what Putnam originally meant. Putnam (1975, p. 200) suggests that a reference-fixing event consists of perceptual contact with some phenomena and a causal description singling out the referent as the thing that is responsible for the phenomena in a certain way. Putnam argues that even if the causal description turns out to be false, and if there is "a good reason to treat it as a mis-description… (rather than a description of nothing at all)", the reference should still be regarded as successful (*ibid*., p. 201).

Along the same line, Putnam (1978, p. 23) offers a methodological maxim, "the principle of the benefit of doubt," which dictates that "when speakers specify a referent for a term they use by a description and, because of mistaken factual beliefs that these speakers have, that description fails to refer, we should assume that they would accept reasonable reformulations of their description." Putnam discusses the example of Bohr's "electron". Although Bohr's description associated with electrons turns out to be false, since Bohr would have accepted reformulations of his description of electrons and indeed did in his later work, we should regard Bohr's use of "electron" as a successful reference. However, Putnam (*ibid.*, p. 25) explicitly argues that 'phlogiston' fails to refer to oxygen even if the principle of the benefit of doubt is applied. According to Putnam, for 'phlogiston' to refer to oxygen, the reformulation of the description involved would be "unreasonable." The descriptions associated with 'phlogiston' are not even approximately true, in contrast to the case of Bohr's "electron."

Therefore, according to Putnam, the theory-laden causal description involved in the reference-fixing event does play a role in whether the reference is successful or not. Not all false causal descriptions can yield a successful reference fixing. The causal description needs to be approximately true so that "there is a good reason to treat it as a mis-description" and a reformulation is "reasonable."

I think Putnam proposes a genuine causal descriptivist account for unobservable kinds. The problem is not that it fails to account for the referential failure in cases such as 'phlogiston' but that it fails to provide a satisfying *criterion* that distinguishes cases of referential success from cases of referential failure. It is simply unclear what makes a causal description a "mis-description" instead of a "description of nothing at all," and what makes a reformulation of a causal description "reasonable" instead of "unreasonable." Putnam's account is not purely causal because the causal description and the underlying theoretical content do matter, but it suffers from a demarcation problem distinguishing approximately true descriptions from completely false descriptions.

Appealing to the empirical success of the theory from which the description is derived cannot solve the problem. Laudan (1981) famously and correctly points out that the history of science offers many examples of empirically successful theories whose central theoretical terms are non-referential. The phlogiston theory is one of them. If we regard all empirically successful theories as approximately true so that a reasonable reformulation of the causal description associated with the central theoretical terms is permitted, 'phlogiston' would refer to oxygen.

Cummiskey (1992) suggests that whether a description is approximately true depends on whether the research program proves to be fruitful in subsequent scientific developments: "If the rough, tentative, programmatic, inductively open-ended specification is further developed and articulated by the subsequent history of a successful theory using the term in question, then we are surely justified in believing that the term has indeed locked onto a causally and explanatorily significant feature of the world" (p. 35-36). Iikawa and Sasaki (2024), following Psillos (1999, Ch. 12), propose a more concrete account appealing to future theories. They suggest that (1) "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*;" and (2) there needs to be a set of descriptions *D* that both past speakers and a relevant future theory associate with *t*, and the core causal description associated with *t* in the future theory needs to explain why *D* denotes *x* (p. 8). The idea is that the early speakers of a scientific term do not need to get everything right to make a successful reference. In fact, Iikawa and Sasaki suggest that the core causal description associated with a term, denoting the kind-constitutive properties, can change as science advances, in contrast to Psillos' (1999) original idea that a term only refers if the core causal description associated with the term is true. According to Iikawa and Sasaki, as long as there is a set of descriptions that survives theory change and can be explained by the future core causal description, the early use of the term does successfully refer.

I agree that Psillos' (1999) requirement that the core causal description, denoting the kind-constitutive properties, be true is too strict, as Psillos himself later acknowledges (2012). A key difference between Iikawa and Sasaki's account and Psillos'—and the reason I prefer to follow the latter—is that the former lacks a causal component linking the use of a scientific term to observed phenomena in the empirical success of a scientific theory. Since defenses of scientific realism typically emphasize that the empirical success of a scientific theory justifies a positive epistemic attitude toward its unobservable entities, it follows that empirical success should also ground the reference of theoretical terms. Throughout scientific development, scientists sometimes discover new entities through experiments while the underlying theories are later superseded, and at other times propose theoretical hypotheses that are later empirically confirmed. Iikawa and Sasaki's account does not capture this distinction, and I will show how this can be achieved in Section 5.

In the rest of the paper, I propose a causal descriptivist account in which the causal description in the reference-fixing mechanism is *exactly true* and does not need to denote the kind-constitutive properties of a theoretical kind. The biggest concern is that such an account would face the same problem as classical descriptivism, namely that it fails to establish any referential continuity across theory changes (Hoefer and Martí 2020). I show that this can be resolved by including an operational description of the specific experimental setups and contexts.

3 Searching for the Exactly True Description

Let me start by considering Psillos' (2012, p. 222) account of the reference-fixing mechanism:

R(x) = x causes phenomena ϕ and D(x).
Term t refers to x if and only if R(x).

Where D(x) is "a description of the ways in which the posited referent [x] is supposed to be causally connected with the phenomena." According to Psillos, 'phlogiston' could not have been fixed to refer to oxygen because the causal role attributed to phlogiston in bringing about combustion is different from the causal role oxygen plays in combustion.

However, the "most serious" problem with this account, as Psillos himself recognizes, is what the size of the causal description should be. On the one hand, if the causal description is too rich, "even though it may explain in virtue of what putative mechanisms [the referent] is supposed to play its causal role, it will be difficult to ensure that it is satisfied by anything, let alone to guarantee some referential continuity in theory change" (p. 224). On the other hand, if the causal description is too thin, it again faces the problem of making reference too easy or failing to pick out a unique referent. Psillos' answer to this challenge appeals to the linguistic behavior of the scientific community. He argues that "care is generally taken [by scientists] to associate with a term enough descriptions to pick out its referent uniquely and to ground its putative explanatory role, but not too many to forestall any referential continuity" (p. 225).

It is unclear how this argument can apply to all cases of referential continuity that scientific realists want to preserve. Take 'electron' as an example. During the early experimental interaction with electrons, the properties that scientists firmly associated with the kind included having a definite position and momentum at all times, having a fixed mass, and having a fixed charge $-q$. As quantum theories develop, most of these properties turn out not to be possessed or stably possessed by electrons. Having a definite position and momentum at all times is shown to be false by the Heisenberg Uncertainty Principle. Having a fixed mass and charge also becomes controversial due to the need for renormalization methods in empirically successful relativistic quantum field theory (Wallace 2021). Even being strictly localizable and countable, which is assumed in non-relativistic quantum mechanics, turns out to be false in the relativistic regime (Teller 1995; Malament 1996; Halvorson and Clifton 2002; Fraser 2008).

As a result, we cannot simply rely on scientists' carefulness in linguistic behavior, hoping that they will always use descriptions that can single out the referents while maximizing the possibility for referential continuity. To understand how referential success and continuity can be established, consider the example of Priestley's 'dephlogisticated air.' In 1774, Joseph Priestley focused sunlight on a sample of mercuric oxide and found out that the gas emitted was "five or six times better than common air for the purpose of respiration, inflammation, and I believe, every other use of common atmospherical air" (Priestley 1775, p.387). Priestley called this gas "dephlogisticated air," based on the phlogiston theory that the gas supported combustion better because it had no phlogiston in it and thus could absorb phlogiston better. From the perspective of contemporary chemistry, Priestley's experiments successfully isolated oxygen, but he described it falsely. Did Priestley succeed in referring to oxygen when he pointed to the jar of gas he managed to isolate? I think he did.[4] The referential success of Priestley's 'dephlogisticated air' is largely grounded in

[4] Kitcher (1993) suggests that there were two modes of reference failed to refer to anything when he intended to refer to the gas that fits the description according to the phlogiston theory. Here, I only consider the first mode of

his operational success of isolating oxygen instead of his linguistic behavior or theoretical beliefs. The referential continuity from ‘dephlogisticated air’ in Priestley’s language to ‘oxygen’ in today’s language is supported by the fact that, according to contemporary chemistry, we can still use Priestley’s method to isolate oxygen.

Similarly, according to our best current theory of electrons, i.e., the Standard Model, the fact that an early cloud chamber can causally interact with electrons still remains true. In all cases of referential continuity, we can safely say that **there is a continuity of experimental success that we have been producing, manipulating, detecting, measuring, or causally interacting with the same kind of things**. But we need to be careful that referential success cannot be equivalent to operational success, which faces the danger of collapsing scientific realism into operationalism. The explanatory success of the superseded theories in revealing the *underlying causal structure* of the specific experiment is also crucial to referential success. This analysis aligns well with Chakravartty's (2007, Sec 2.4) strategy against the pessimistic meta-induction. He classifies theoretical properties of a putative entity into two categories: detection properties, which are causally linked to experimental apparatus, and auxiliary properties, which are other putative properties postulated by the theory. The distinction is epistemic: based on empirical success, a realist can commit to detection properties and remain agnostic about auxiliary properties. As a result, the causal description in the reference-fixing mechanism should capture the detection properties of the referent, which is responsible for the continuity of experimental success.

However, detection properties alone are not sufficient to make up the causal description for reference fixing. Recall the example of 'electrons' above. In early experiments of detecting electrons, a detection property attributed to electrons was having a definite position at all times. This was reasonably postulated by scientists' (theoretical and experimental) understanding of the causal structure of a particle detector, such as a cloud chamber, from which the particle trajectory can be reconstructed. However, this property is later falsified by quantum mechanics and quantum field theory. If the early reference of 'electrons' was fixed through detection properties only, we can hardly maintain that early scientists indeed referred to the same elements of reality as we do today.

Therefore, more contextual details about the experimental success need to be included along with the detection properties in the causal description. Nola (1980) proposes a causal descriptivist account along this line. He suggests that when scientists introduce the name *T* for a putative kind, they will form a belief about *T* of the form: “All *T* have [causal] power *P* which in circumstances *C* cause [phenomena] *O*” (p.524). According to Nola, the causal description in the reference-fixing mechanism should consist of specific causal powers attributed to the referent by the theories, explaining how the referent causes the relevant phenomena under specific conditions. As a result, the causal description does not need to consist of some essential or general properties of the referent but can consist of contingent, context-dependent properties of the referent specific to a certain experimental setup.

---

reference. I will discuss Kitcher’s analysis in more detail in Section 5. when Priestley used the term ‘dephlogisticated air.’ He succeeded in referring to oxygen when he intended to refer to the gas isolated in his experiment, and he

The inclusion of circumstances C in the causal description is crucial. To see how this can help provide an exactly true description for a reference-fixing mechanism, consider the early theoretical belief about electrons having definite position and momentum again. Under the specific conditions of the early experiments on electrons, electrons indeed can be described as having definite position and momentum. The restriction from the Heisenberg Uncertainty Principle, $\Delta x \Delta p \geq \hbar/2$, does not contradict such a description because the scale of the experimental resolution of, say, a cloud chamber is larger than the scale of $\hbar$. Therefore, within the experimental limitations of the early apparatus, electrons indeed *appear to* have definite position and momentum.

My use of “appearing to” here needs further clarification. The theoretical description of electrons at the most fundamental level currently available to us is provided by quantum electrodynamics (QED). According to QED, electrons do not have definite position and momentum. However, even in QED, electrons can still be approximately localized in a region so that particle trajectories can be constructed in experiments (Bain 2000). In other words, within the restricted domain of the experimental apparatus, QED can be reduced to a higher-level electron theory where electrons can be described as having definite position and momentum. What is reduced in this process is the effects that are physically negligible within the domain. Neglecting these effects does not harm the explanatory power of the higher-level description, which gives a complete description of the physically significant causal structure. Therefore, when I said that electrons “appear to” have definite position and momentum, I meant that although electrons do not possess these properties according to the more fundamental description, electrons do possess these properties according to a higher-level description where some effects are reasonably neglected.

In general, as long as we can specify the context and experimental limitations in the causal description, the later falsified causal properties attributed to a referent can indeed appear to be possessed by the referent and help fix the reference. However, how the contexts and specific conditions can be specified needs further examination. In most cases, the empirical limitation and applicable domain of a theory are only specifiable in the successive theory. For instance, only after the development of relativity theory can we understand the fact that Newtonian mechanics is empirically accurate within the “classical limits.”

Rivat (2025) proposes a solution using the framework of effective theories. According to Rivat, the causal description in the reference-fixing mechanism is indexed to a limited range of parameters over which the theoretical description is empirically accurate. While the speaker is not required to know the exact value of the range of parameters specifying the applicable domain, Rivat suggests that the empirical inputs and the effective theory methods allow the speaker to "estimate the range over which the effective law is likely to remain empirically accurate" (p. 17). A causal description used to fix a reference is thus required to be exactly true when indexed to a range of parameters specifying the applicable domain.

I agree with Rivat that in most cases of referential success throughout scientific development, the underlying theory can be regarded as an effective theory that is empirically accurate within a restricted domain. As a result, Rivat's treatment can account for the referential continuity of most scientific terms, such as 'electrons,' across theory changes. However, Rivat's strategy cannot account for cases in which a reference is successful even though the underlying theory and dynamical law are completely wrong, and thus cannot be regarded as an effective theory. "Dephlogisticated air," in Priestley's experiment of isolating oxygen, would be one example. Therefore, I propose a more general account that preserves Rivat's success but does not rely on the effective theory framework.

The key to my account is to realize that scientists during a period of scientific development do have epistemic access to the experimental contexts. What they do not have epistemic access to is how future theories would characterize their theory's applicable domain, but they are aware of the contexts and conditions under which their experiments have been successful. I suggest that the domain specification involved in the reference fixing should not be a description of *what-is,* but a description of *how-to*. What the scientists in the early 1900s did know was how to build a cloud chamber. They had perfect access to knowledge about all the engineering and error-testing details in order to make a working cloud chamber. And the fact that a cloud chamber can causally interact with electrons remains exactly true in the successive theories. According to the Standard Model, we can still build a good old cloud chamber according to the instructions offered by the early scientists, and the cloud chamber we build will still detect electrons.

In Rivat's account, the applicable domain of the detection properties is estimated within the underlying theory based on its empirical success, while in my account, the applicable domain is specified by describing the details of how to build the apparatus (e.g., 'pour some isopropyl alcohol into a clear-sided container…'), and in what environmental contexts the experiments are to be performed (e.g., 'at room temperature and pressure'). In this way, the theoretical estimate of the applicable domain is replaced by precise operational and contextual knowledge, while the success of Rivat's account is preserved. The advantage of my account is that it is not bound to the effective theory framework and thus can apply to cases even when the framework is not applicable. For instance, when Priestley (1775, pp. 387-8) describes the air he isolated, the underlying theoretical description, "the fitness of air for respiration depends upon its capacity to receive the phlogiston exhaled from the lungs," is completely wrong and has no applicable domain. Yet the causal properties he associates with the air, such as being "better…for the purpose of respiration, inflammation," "a candle burned in this air with an amazing strength of flame," and the operational knowledge of how he isolated it, together establish a successful reference to oxygen.

Note that I do not mean to classify scientific knowledge into theoretical and experimental knowledge. The distinction at concern here is purely pragmatic, based on the role a piece of knowledge or information plays in a specific reference-fixing mechanism. For instance, when we describe the environmental context of an experiment in particle physics, we might say, 'the temperature is around 20 degrees Celsius.' The terms 'temperature' and 'Celsius' are theory-laden, but the description enters the reference-fixing mechanism for a particle term as contextual knowledge.

In summary, following Psillos' account above, I suggest that the causal description D(x) in the reference-fixing mechanism should consist of two parts:

a) the causal (or detection) properties attributed to the referents in bringing about the relevant phenomena;

b) the detailed experimental setups, including operational knowledge and environmental contexts, under which the referents appear to have these causal properties.

Part a) preserves the causal description in Psillos' original proposal: "they are meant to offer enough identifying markers of an entity (related to its causal role vis-à-vis phenomena Φ) to allow the stable use of the term in certain inductive and explanatory practices" (2012, p. 224). Although Psillos does not make a distinction between reference to single individuals and reference to kinds, I shall take the liberty of improving the clarity here. When scientists intend to refer to a token unobservable entity (e.g., a specific black hole, 'Messier 87'), part a) needs to offer enough identifying markers to single out that specific individual in the experimental context. When scientists intend to refer to a theoretical kind, part a) needs to offer enough identifying markers of the specific kind in the experimental context, which help avoid the qua-problem faced by the pure causal theory of reference for natural kinds (Devitt and Sterelny 1987, sec. 5.3). The qua-problem, simply put, is that when we point to a sample hoping to fix the reference of a kind term, it is unclear which kind we are actually referring to among all the kinds that the sample instantiates. Devitt and Sterelny, as well as Stanford and Kitcher (2000), suggest that the term introducer needs to associate some description with the kind term that determines the specific kind she intends to refer to.

Part a) serves this role for theoretical kind terms. For a reference to be successful, part a) needs to provide sufficient causal properties to identify a unique theoretical kind such that *all and only* the members of the kind possess all of them in the specific experimental context. Note that the identifying markers of a theoretical kind do not need to be the kind-constitutive properties (or the essential properties of a kind, if one commits to essence kinds), as Psillos (2012) revises his (1999) account. Here, I allow more explicitly that it can be contingent and context-dependent properties, which aligns with Stanford and Kitcher's solution to the qua problem. For instance, when J. J. Thomson (1897) discusses the "corpuscles" in cathode rays, the causal properties he associates with the term include being "charged with negative electricity" and having a ~1400 times smaller mass-to-charge ratio than that of a Hydrogen atom, which provide sufficient identifying markers of electrons in the experimental context. In Priestley's experiment of isolating oxygen, even if the sample that he isolated was impure, the causal description Priestley associated with it, such as being "produced from *mercurius calcinatus per se*" and being "better…for the purpose of respiration, inflammation," provides sufficient identifying markers of what we call oxygen today. Priestley did not know the underlying kind-constitutive properties of oxygen at that time. What licensed Priestley to refer to oxygen is that the set of entities singled out by his context-dependent causal description is *identical* to the set of entities singled out by our description of oxygen, while the descriptions used for constructing the two sets can be different.

Moreover, part a) ensures that this account does not collapse into operationalism. Chang (2011) argues that operational stability is what is responsible for fixing the extension or reference of a scientific term. Since phlogiston theory is associated with a set of repeatable and robust phenomena, one might think that 'phlogiston' enjoys the same kind of operational stability as well. The difference in referential status between ‘electrons’ in the early 1900s and ‘phlogiston’ is made clear by part a) of the causal description. The causal properties attributed to electrons in the early 1900s, such as having definite position and momentum, having fixed charge and mass, are indeed possessed by electrons under the experimental setups at that time. The causal properties attributed to phlogiston, such as being a component of a combustible substance and being released from the substance, causing combustion, are not possessed by anything in reality, even if the experimental conditions are perfectly specified.

At the beginning of the introduction of a term, a large portion of b) might very well be over-specified. For instance, it might be that the first cloud chambers that have been built are located in the northern hemisphere, so 'being in the northern hemisphere' will be included in part b). The reason is that the term introducer is required to know both part a) and part b) for the reference-fixing event to succeed, but she might not be able to identify the relevant contextual information and leave out the irrelevant ones. But that is fine. In a how-to description, it’s better to be careful instead of sloppy. One might worry that the term 'electrons' is then fixed to refer to the electrons in the northern hemisphere only. This worry will be properly addressed in the next section. I will show that over-specification and context-dependency of the causal description do not affect the meaning of the term, which consists of the referent only.

The speaker also does not need to explicitly utter everything relevant to part b) during a reference-fixing event. A scientist can simply point to an experimental apparatus, and such ostension can serve the purpose of part b) as long as the scientist is cognitively aware of the engineering details of the apparatus and the conditions under which it has been working.

In summary, I have shown that the causal description in the causal descriptivist theory can be exactly true and survive through theory change once the experimental contexts and conditions are specified, and scientists at a time are indeed cognitively aware of these contexts. In the next two sections, I will further furnish the account by considering two possible challenges. I will show that the causal descriptivist theory, under my analysis, can adequately account for scientists’ complex linguistic behaviors in concord with scientific realism.

4 The Meaning of Causal Descriptivist Terms: The Externalist Story

In this section, I will consider a challenge to the causal descriptivist account concerning the meaning of a term fixed by a causal description. Nola discusses a peculiar result of his account: if a specific causal power, *P*, is involved in the reference fixing of a term, *T*, then ‘*T* possesses *P*’ becomes an *a priori* truth. This result is undesirable. First, it defeats the purpose of the hybrid account being genuinely causal. One of the main advantages of Kripke and Putnam’s proposal is that subsequent speakers can successfully refer to the referent as long as their usage is grounded in a causal chain of reference borrowing, even though the speakers might be ignorant of any description associated with the term. A later speaker should be able to refer to the referent by *T* without knowing anything about *P*. Second, the meaning of the term becomes very narrow. It is highly unlikely that every time contemporary scientists use a scientific term, they have in mind the early theory and experimental setup that originally fixed the reference. Co-reference also becomes difficult to achieve. For instance, scientists found out that 'beta particles' emitted during the beta decay process are just 'electrons' after the reference for both terms had already been fixed. But if the meanings of both terms were fixed through different experimental setups, in what sense can we say that they co-refer to the same entity?

The same kind of worry motivates Psillos (1999, Ch. 12) to suggest that the core part of the causal description in the reference-fixing mechanism needs to be about the essential properties of the kind.[5] It is crucial in Putnam and Kripke's original causal account that manifest properties cannot fix the reference of a natural kind. To preserve this feature, Psillos (1999) believes that if the causal description only includes contingent properties of the referent, the reference cannot be fixed.

Similarly, Rivat (2025, p. 20) suggests that "look first for unobservable entities that fall within the empirical reach of the theory and assume that its term fails to refer to anything real in unexplored domains until there is evidence to the contrary." Rivat seems to suggest that since the causal description is indexed to a specific domain, the resulting referential success and ontological commitment also become indexed to a specific domain. But if we index the referential terms to a specific domain, in what sense can we say that 'electrons$_{\text{low-energy}}$' and 'electrons$_{\text{high-energy}}$' co-refer to the same thing?

The core assumption underlying these worries is that the causal description involved in the reference-fixing mechanism becomes part of the *meaning* of the term. I imagine the reasoning to be something like this. In the pure causal theory of reference, the reference-fixing event does not involve any description that is required to be exactly true of the referent. The causal link between the referent and the introduction of the term can be established without any description, and the meaning of the term only consists of the referent. In a causal descriptivist account, since the causal description enters as part of the reference-fixing mechanism, the resulting meaning of the term consists of both the referent and the causal description.

This reasoning, however, is mistaken. In the following, I will argue that the causal description, although being part of the reference-fixing mechanism, does not become part of the meaning of the term. My strategy is to return to the most uncontroversial cases of the causal theory of reference, namely the reference fixing of observational proper names, locate a component there, and show that this component plays the same role as the causal description does in the reference fixing of theoretical terms. If the causal theory of reference can fix the reference of an observational proper name to the referent only, then a causal descriptivist account can also fix the reference of a theoretical term to the referent only.

Consider a paradigm case of the causal theory of reference, the reference fixing of an observational proper name. Following Devitt (1981, p. 26), the paradigm reference-fixing event of a proper name is when the speaker is in a face-to-face confrontation with the object, say, a cat, and

[5] Psillos later in his (2012) withdraws this requirement and suggests that contingent properties are sufficient to fix the reference, as long as they provide enough identifying marks to pick out the referent uniquely.

says, 'Let's call her Nana.' The speaker's ability to designate the cat and fix the reference in this case is made possible by her perception and mental representation of the cat. It is really *this* cat that the speaker has in mind when she names her.

Now, consider a different scenario. Suppose some evil technology alters the neural system of the speaker, and the transmission of signals from her eyes to her brain is intercepted. There is in fact a cat in front of her. However, the evil technology feeds her brain an image of a different cat. Without knowing anything about the evil technology, the speaker says, 'Let's call her Nana.' Clearly, in this case, the speaker fails to fix the reference to the real cat in front of her. Without an actual perception and mental representation of the cat, a causal link is not established between the cat and the name. The speaker does not have the ability to designate the cat.

The upshot of the two examples is this. As Devitt points out, a speaker's ability to designate and name an object relies on the speaker *having the object in mind*. In the observable cases, even though no description is involved in the reference fixing, a mental representation grounded in the object is nevertheless required, ensuring that the speaker indeed has the object that she intends to designate in mind. This is essentially Putnam's (1981) argument against what he calls the magical theories of reference. Here's an example that Putnam gives: "A discourse on paper might seem to be a perfect description of trees, but if it was produced by monkeys randomly hitting keys on a typewriter for millions of years, then the words do not refer to anything." (p. 4) Words only refer if the speaker has the ability and intention to refer.

Putnam's view is often termed semantic externalism, which I will extend to the causal descriptivist account in the following. In order for one to meaningfully use a term to refer to something in reality, one's usage of the term needs to be properly grounded in a causal chain linking back to the referent. During a reference-fixing event, being the beginning of a causal chain for the usage of a term, the speaker thus needs to be in a proper causal relation with the referent. In the paradigm cases, this causal link is established by the speaker having a mental representation grounded in the referent.

Let's now return to the unobservable cases and see what allows a speaker to have the referent in mind, thus possessing the ability to designate. If the entity that the speaker tries to name is unobservable, the speaker only has a mental representation of the phenomenon caused by the entity. If the speaker only has the mental representation of the phenomenon and the thought "Whatever causes this phenomenon," we can hardly say that the speaker has *the object* in mind. Consider an example similar to what Enć (1976) discusses. Suppose Smith, who owns a garden, notices one morning that his lettuces are destroyed and says, "I'll call whatever destroyed my lettuces Snowwhite." The next morning, Smith discovers that his teacup is destroyed and says, "Snowwhite destroyed my teacup," or, "Snowwhite did not destroy my teacup." What prompts Smith to reasonably assert either? As Enć points out, in order for Smith to designate the object that destroyed his lettuces in later sentences, Smith must have at least inferred some causal properties that the object has, for instance, from the teeth marks on the lettuces, from the footprints on the floor, and so on, when he introduces the name. Without any of these thoughts, Smith does not have the ability to designate the object by the name "Snowwhite."

Therefore, for unobservable entities, since the speaker cannot have a mental representation of the referent, the speaker needs to have a causal description of the referent in mind in order to gain the ability to designate the object. The causal link between the speaker's use of a term and the referent is established by (1) the causal link between the speaker and the phenomena, (2) the causal link between the phenomena and the referent, and (3) the speaker's epistemic access to (2). The mental representation of the phenomenon plus the causal description in unobservable cases thus plays the same role as mental representation in the observable cases. They ensure a causal link is established between the entity and the speaker's use of the term by ensuring that the speaker indeed has the entity in mind, thus granting the speaker the ability to designate.

For theoretical kind terms, a causal link between the samples and the speaker's use of the term is not sufficient to grant the speaker the ability to designate the kind because of the qua-problem. The speaker also needs to have in mind which specific kind, among all that the samples instantiate, that she intends to refer to. Therefore, the causal link between the speaker's use of a kind term and the referent kind is established by (1) the causal link between the speaker and the sample phenomena, (2) the causal link between the sample phenomena and the sample entities, (3) a necessary and sufficient condition—possibly context-dependent and nonunique—for the referent kind that the sample entities belong to, and (4) the speaker's epistemic access to (1), (2) and (3). The causal description D(x) in the reference-fixing mechanism, as I discussed in the previous section, needs to establish both (2) and (3) for the reference to be successful.

One might object to this analogy by arguing that in the paradigm cases, the mental representation of the referent can be inaccurate, yet in the unobservable cases, the causal description is required to be exactly true. Consider Kripke's (1980, p. 118) example that gold might not be yellow. "Suppose an optical illusion were prevalent, due to peculiar properties of the atmosphere," which caused us to have a mental representation of gold as yellow metal, but, in fact, gold was blue. Kripke argues that in this case, we would still have successfully designated gold, even though our mental representation was wrong. However, it is salient in this example (in contrast to my example of evil neuro-technology above) that our false mental representation is indeed partially caused by, or in Devitt's terms, grounded in the thing, i.e., the samples of gold. There must be a contextual explanation of how the blue thing appeared to be yellow in our eyes, which in this scenario is the peculiar properties of the atmosphere. Even though we might not have epistemic access to this fact, we have epistemic access to the contexts of our perception, such as where and when we perceived the samples of gold to be yellow. Therefore, this case, in contrast, shows that mental representation in the observable cases is indeed very similar to the causal description in the unobservable cases. Manifest properties in the mental representation can be regarded as exactly true if the contexts are specified in the form "the thing x appears to be P under circumstances C."

To conclude, if mental representation enables a speaker to name an object so that the meaning of the name consists of the object only in the observable cases, causal descriptions plus mental representation of the relevant phenomena should also enable a speaker to name an object so that the meaning of the name consists of the object only in the unobservable cases. The causal description serves as an auxiliary mediator that establishes the causal link between the object and the speaker's use of the term, granting the speaker the ability to designate the object. As long as the speaker has the ability to designate the object at the reference-fixing event, the meaning of the term becomes fixed to the object only. Although the speaker at the reference-fixing event needs to have the causal description in mind, later speakers who borrow the reference are not required to as long as their usages are grounded in the causal chain linking back to the reference-fixing event.

If the meaning of a scientific term, once fixed through a causal description in relation to certain phenomena, consists of the referent alone, we do not need to accept Nola's peculiar result that the causal description in the reference-fixing mechanism is *a priori* true of the referent. This way of understanding the reference-fixing mechanism also allows us to understand scientists' complex linguistic behaviors in scientific development in a better way.

First, after scientists discover the existence of an unobservable kind through some specific experimental setups and fix the reference for a scientific term, they continue to discover more causal properties of the entities and refine our theoretical understanding of them. Since the meaning of the term only consists of the referents but not the causal descriptions initially used to introduce the term, we do not need to regard later scientific development as "refining the reference," as Psillos (2012) suggests. The meaning of the term has not changed. We are simply conjecturing, testing, and learning new properties (possibly in different domains) about the same things. The meaning of the term is also not narrowly tied to early experimental setups, allowing later scientists to designate the referents freely without having the early experimental setups in mind.

Second, two terms can co-refer to the same entities even if they are introduced through different causal descriptions and phenomena, and even when the speakers are not aware of this. For instance, when beta decay was first studied in experiments, the emitted particles were named 'beta particles.' The causal description fixing the reference for the term thus consists of the referents' causal role in the beta decay process and the description of relevant experimental setups. We now know that 'beta particles' are just electrons, but early scientists were not aware of this. The meanings of 'beta particles' and 'electrons' have been identical all along. The statement 'beta particles are electrons' is an example of what Kripke (1980) regards as an *a posteriori* necessary truth. 'Beta particles' and 'electrons' designate the same entities, but it requires theoretical and empirical work to find out about this identity relation. Similarly, Priestley's use of "dephlogisticated air" during his experiment and the term 'oxygen' as we use it today co-refer to the same entities without Priestley's awareness.

Similarly, the same term can refer to different entities if it is introduced through different causal descriptions and phenomena. For instance, Pauli in 1930 introduced the term 'neutron' to refer to the undetected particles emitted during the beta decay process. Chadwick (1932) also used the term 'neutron' to refer to the massive, neutral particles he discovered in his experiments. According to the Standard Model today, Pauli referred to neutrinos, and Chadwick referred to neutrons. The externalist picture can account for this scenario with no problem. Although using the same term, the two scientists referred to different entities in reality because they had different causal descriptions in mind. A parallel example in daily life is when two people have the same name. Which person a speaker refers to by uttering the name depends on which person the speaker has in mind.

In summary, the causal description in the reference-fixing mechanism does not enter as part of the meaning of the term but should be regarded as an auxiliary requirement for the speaker to have the referent in mind, just like the role mental representation plays in paradigm cases of the causal theory. The contingency and context-dependency of the causal description thus do not affect the meaning of the term, which consists of the referent only.

5 Theoretical Terms Before Reference Fixing

Another challenge to the causal descriptivist account is that it fails to account for the reference of scientific terms that are not introduced by being associated with any phenomenon. Papineau (1996, p.4) expresses this worry: "The causal theory seems unable to account for terms, like 'positron,' 'neutrino,' and 'quark,' that are explicitly introduced to refer to hypothetical entities which are conjectured to play certain theoretically specified roles, before any direct experimental manifestation of these entities is available for any dubbing ceremony."

Hoefer & Martí (2020) present a similar argument. They use the theoretical development of the neutrino as an example. They argue that "neutrinos were not introduced as the cause of any phenomenon φ; they were introduced to 'balance the books' of conservation principles, being the bearers of the energy and angular momentum that seemed to be missing in beta decay processes" (p.15). Whether the example of neutrinos serves the purpose of the argument might be debatable here. By the time Pauli and Fermi postulated a different particle emitted during beta decay, beta decay processes had been largely studied by various experiments. It could be argued that the introduction of neutrinos was indeed associated with a set of phenomena (Brown 1978).

Nevertheless, I agree that the general worry remains and needs to be accounted for. Consider that when Gell-Mann (1964) first proposed the quark model based on "the eightfold way" scheme, the proposal was not strongly associated with any particular experimental phenomena. It is clear that, according to the causal descriptivist account, Gell-Mann's theoretical development cannot causally fix the reference of 'quarks' to quarks in reality.

However, this result does not imply that all scientific terms whose references are not causally fixed turn out to be meaningless. Once we understand that the causal description serves the same role as mental representation in the paradigm cases of the purely causal theory, we can utilize the tools that the causal theorists have already developed. The specific solution lies in Donnellan's (1966) distinction between referential use and attributive use of definite description, which is widely adopted by causal theorists (Devitt 1981). Consider the following example presented by Donnellan:

> "Suppose one is at a party and, seeing an interesting-looking person holding a martini glass, one asks, 'Who is the man drinking a martini?' If it should turn out that there is only water in the glass, one has nevertheless asked a question about a particular person, a question that

it is possible for someone to answer. Contrast this with the use of the same question by the chairman of the local Teetotalers Union. He has just been informed that a man is drinking a martini at their annual party. He responds by asking his informant, 'Who is the man drinking a martini?' In asking the question the chairman does not have some particular person in mind about whom he asks the question; if no one is drinking a martini, if the information is wrong, no person can be singled out as the person about whom the question was asked." (p. 287)

According to Donnellan, the definite description, "the man drinking a martini," is used referentially in the first case and used attributively in the second case. The difference is whether the speaker indeed has the referent *in mind*. The first speaker has a perception and mental representation of the man she intends to refer to, which is grounded in the referent, enabling her to designate the man. The second speaker, in contrast, does not have a perception and mental representation of the man he intends to refer to, and his use of the phrase is also not grounded in a causal network of reference borrowing linking back to a reference-fixing event.

Because of this difference, how the truth value of the description can affect referential success is different. In the first case, when a description is used referentially, the speaker's description and mental representation can be false of the referent but still achieve referential success. Although the speaker misdescribes the man she has in mind, her description and mental representation are nevertheless grounded in the man. As I argued in the previous section, her description and mental representation are true of the referent in the sense that the man indeed appears to be drinking a martini in her perception. In the second case, when a description is used attributively, the speaker cannot achieve referential success if nothing fits the description. The underlying semantics is descriptivist, as Russell (1905) originally suggests. The description, "the man drinking a martini," successfully denotes someone if there is a person at the party who is a man and drinking a martini. The description is denotation-less if no one at the party is a man and drinking a martini.[6]

Now, let's return to theoretical descriptions in scientific discourse. A theoretical description can be used referentially if it enters the reference-fixing mechanism as a causal description. Such a theoretical description is used referentially because the speaker has the referent in mind and designates it. If a theoretical description is not used as a causal explanation of some phenomena, it should be regarded as being used attributively. The term 'quarks' in 1964 successfully denoted something if there is indeed something in reality that fits the theoretical description of it. The term was denotation-less if there is nothing in reality that fits the description. In either case, the term was meaningful. The meaning of the term, whether or not it denoted something, was the theoretical description associated with the term. This semantic analysis coheres with scientists' beliefs about whether a theoretical term refers before it is connected to any empirical phenomenon. If a scientist

[6] A note on the terminology here. Following Russell, a term is said to denote something as long as it refers to the thing. The underlying semantics for the successful reference can be either descriptivist or causal. A term is said to designate something if the reference of the term is *fixed* to refer to the thing, according to the causal theory of reference.

in 1964 believed that there was something in reality that ‘quark’ referred to, her reason must be that there is something in reality that fits the theoretical description of ‘quarks.’

What happened when scientists first empirically confirmed the quark model and realized that ‘quarks’ indeed existed (Breidenbach et al. 1969)? Consider Donnellan’s example again. Suppose the chairman of the local Teetotalers Union asks his informant to bring the man drinking a martini to him, and in fact, there is a man at the party who is drinking a martini. After the informant brings that man to the chairman, the chairman says, “A-ha, *you* are the man drinking a martini!” During this face-to-face confrontation, the chairman forms a mental representation of the man whom the attributive description denoted. At a later time, the chairman says to others, “I should charge the man drinking a martini a fine!” He is now using the description referentially, for he has a mental representation of the man he intends to refer.

As a result, an attributive description can turn into a referential description through a reference-fixing event. In other words, the reference-fixing event might not occur when a term is first introduced, but can still occur at a later time after the term has been used attributively. After we have connected Gell-Mann’s quark model to empirical evidence through a causal description, the term ‘quarks’ is now fixed to refer to quarks in reality.

One might worry that if the attributively used definite description is required to be exactly true to successfully denote something in reality, many usages of scientific terms in the early stages before reference fixing could turn out to be denotation-less. First, this result is perfectly consistent with scientific realism. The standard interpretation of the No-Miracles argument, namely that it would be a miracle if our scientific theory achieved enormous empirical success while failing to describe reality accurately, only argues for a positive epistemic attitude toward the theories and the postulation of theoretical entities that have been *empirically successful*. If a piece of theoretical content has not yet been connected to any observable phenomenon, a scientific realist is not justified in committing to the putative entities it postulates, regardless of whether the piece turns out to be true or not.[7] As a result, my analysis of the referential status of a scientific term largely coheres with the epistemic attitude of a scientific realist such that, whenever a realist is justified in committing to a theoretical entity, the corresponding theoretical term does refer.

Second, being denotation-less does not mean being meaningless. Let's consider Gell-Mann's use of 'quarks' in 1964 again. It is rather debatable whether Gell-Mann's quark model in 1964 provided an exactly true (albeit partial) description of reality and successfully denoted the quarks that we know of today. For the sake of the argument, let us suppose that it did not and examine the semantic consequence in scientific discourse.

In 1965, a scientist could state, 'if quarks were real…' or, 'if Gell-Mann were right about these quarks…" In both cases, the meaning of the term 'quarks' is the theoretical description in Gell-Mann (1964). 'Quarks are real' means that the theoretical description actually corresponds to reality,

[7] This analysis, again, resonates with Chakravartty’s (2007) distinction between detection properties and auxiliary properties. Auxiliary properties are postulated by the theory yet have not been connected with any empirical phenomena. According to Chakravartty, scientific realists can remain agnostic or skeptical about them until they become detection properties as science develops.

which is a false but meaningful sentence. Today, if a scientist states, retrospectively, 'Gell-Mann was right about quarks in 1964,' this sentence could have two slightly different meanings depending on the speaker's intention and the conversational context. One is when the speaker intends to refer to the quarks (as elements of reality) that we know of today, then the sentence means something like, 'Gell-Mann's theoretical description in 1964 truly described this element of reality,' which is false but meaningful. The other one is when the speaker intends to refer to the 'quarks' that Gell-Mann talked about in 1964, then the sentence means something like, 'Gell-Mann's description associated with the term 'quarks' in 1964 corresponds to reality,' which is again false but meaningful.

In summary, all sentences involving 'quarks' are meaningful even if the term is denotation-less. The truth-value of the sentence is determined by whether the description associated with the term corresponds to reality. If one believes that Gell-Mann's (1964) quark model is not exactly true of reality, then one should not take the sentence 'Gell-Mann was right about quarks in 1964' to be true. Alternatively, the sentence 'Gell-Mann was right that there are particles with fractional charge' can be true, since it singles out a part in Gell-Mann's theoretical description that corresponds to reality.

Let's consider the puzzling example of Priestley's "dephlogisticated air" again. Kitcher (1993) suggests that we should distinguish two different modes of reference when scientists talk about a scientific term. On some occasions, by "dephlogisticated air," Priestley and his contemporaries intended to refer to "the kind of stuff that was isolated in the experiments they are reporting," and the term successfully referred to oxygen. On some other occasions, by "dephlogisticated air," Priestley intended that "the referent should be whatever satisfies a particular description", and the term failed to refer (p.102-104).

We can improve on Kitcher's analysis here. Consider Donnellan's example again. Suppose the chairman of the local Teetotalers Union asks his informant to bring the man drinking a martini to him, and the informant brings a random man and makes up a story about how the man was drinking a martini. After the chairman forms a mental representation of the man, he says to others, "I should charge the man drinking a martini a fine!" Now, even though his description of the man is completely false because the man he has in mind was not drinking a martini, he still succeeds in referring to *the man*.[8]

The case of Priestley's "dephlogisticated air" is very similar. I have argued above that causal description, along with mental representation of the relevant phenomena, plays the same role as mental representation in observable cases. When Priestley successfully isolated oxygen in his experiment, he successfully identified the causal properties of oxygen in the specific experimental setup and fixed the reference to oxygen. Although the description implied by the term

[8] Note that the description does not even appear to be true of the referent here because the chairman did not see the man drinking from a martini glass himself. But this example does not contradict what I have argued in the previous section because the false description has nothing to do with the chairman's perception and mental representation, which fix the reference. The false description does not enter the reference-fixing mechanism at all.

‘dephlogisticated air’ is completely false of oxygen, he still succeeded in referring to oxygen because it is the jar of oxygen that he had in mind.

There might be some cases when it becomes ambiguous whether a speaker uses a definite description attributively or referentially. Suppose the chairman of the local Teetotalers Union starts pondering upon details in the informant’s made-up story, such as how the man sneaked the martini into the party, who provided him the martini, and so on. It becomes ambiguous whether the speaker is still referring to the man he has a mental representation of or to a fictional character in the made-up story. This is the same kind of ambiguity we find in Priestley’s use of ‘dephlogisticated air.’ But there are some unambiguous cases that we can establish: before his experimental success of isolating oxygen, he was using the phrase attributively; during his experiments and whenever he was describing his experiments, he was using the phrase referentially. According to Donnellan, whether a definite description is used referentially or attributively is “a function of the speaker’s intentions in a particular case” (p. 297). It is also possible for listeners to misunderstand the speaker’s intention. Therefore, cases of ambiguity can arise when the information to determine the speaker’s intention is insufficient or when the speaker’s intention itself is ambiguous.

One last thing to mention is that, according to Devitt (1981, Ch. 6.5), there is a distinction between designational empty names and attributive empty names. Designational empty names are those grounded in an attempt at reference fixing but fail to refer to anything. ‘Phlogiston’ falls into this category because there was an attempt at fixing its reference through a causal description associated with some phenomena, although nothing fits the causal description. Attributive empty names are used attributively and fail to refer to anything because nothing fits the description. Cases in this category occur when scientists entertain some hypotheses without having any specific phenomenon in mind, and the hypotheses turn out to be false. Causal theorists attempt to maintain such a distinction by providing a specific semantic account for designational empty names (Braun 1993; 2005), which can be applied to cases of scientific terms as well.

6 Conclusion

I have shown above that the causal descriptivist account remains the most viable option in solving the semantic puzzle for scientific realism facing theory changes. The causal description in the reference-fixing mechanism can account for referential failures, and the specification of experimental contexts and conditions in the causal description can provide an exactly true description of the referent within a specific domain, accounting for referential successes. Moreover, once we realize that the meaning of the scientific term fixed through a causal description consists of the referent only, we see that the account has all the advantages regarding reference fixing, reference borrowing, and co-reference that the pure causal theory of reference has, which allows us to account for the complex linguistic behaviors of the scientific community.

Establishing referential continuity in the history of science serves as the first step to defend scientific realism against the meta-induction argument: it shows that not *all* scientific terms in superseded theories failed to refer. The causal descriptive theory refined in this paper provides a semantic framework that accounts for cases in which scientific realists argue that there is genuine accumulation of ontological knowledge. The challenge remains that some theories were once empirically successful, yet their central terms failed to refer. How the refinement proposed in this paper might improve scientific realists' arguments would require future work.

Work Cited

Bain, J. (2000). "Against particle/field duality: asymptotic particle states and interpolating fields in interacting QFT (or: Who's afraid of Haag's theorem?)", *Erkenntnis* 53 (3): 375-406.

Bird, A. (2000). *Thomas Kuhn*, Princeton, New Jersey: Princeton University Press.

Braun, D. (1993). "Empty names", *Noûs* 27: 449-469.

--- --- (2005). "Empty names, fictional names, mythical names", *Noûs* 39: 596-631.

Breidenbach et al. (1969). "Observed behavior of highly inelastic electron-proton scattering", *Physical Review Letters* 23 (16): 935-939.

Brown, L. (1978). "The idea of the neutrino", *Physics Today* 31 (9): 23-28.

Chadwick, J. (1932). "Possible existence of a neutron", Nature 129 (3252): 312.

Chakravartty, A. (2007). *A Metaphysics for Scientific Realism*, Cambridge: Cambridge University Press.

Chang, H. (2011). "The persistence of epistemic objects through scientific change", *Erkenn* 75: 413-429.

Cummiskey, D. (1992). "Reference failure and scientific realism: a response to the meta-induction", *British Journal for the Philosophy of Science* 43: 21-40.

Deutsche, M. (2023). "Is there a '*qua* problem' for a purely causal account of reference grounding?" *Erkenntnis* 88: 1807-1824.

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

--- --- (1996). *Coming to Our Senses: A Naturalistic Program for Semantic Localism*, Cambridge: Cambridge University Press.

Devitt, M. and K. Sterelny (1987). *Language and Reality*, Cambridge, MA: MIT Press.

Donnellan, K. S. (1966). "Reference and definite descriptions", *The Philosophical Review* 75 (3): 281-304.

Enć, B. (1976). "Reference of theoretical terms", *Noûs* 10 (3): 261-282.

Fraser, D. (2008). "The fate of 'particles' in quantum field theories with interactions", *Studies in the History and Philosophy of Modern Physics* 39: 841-59.

Gell-Mann, M. (1964). "A schematic model of baryons and mesons", *Physics Letters* 8 (3): 214-215.

Hoefer, C. and G. Martí (2020). "Realism, reference & perspective", *European Journal for Philosophy of Science* 10: 38.

Iikawa, H. and G. Sasaki (2024). "Temporal externalist descriptivism on natural kind terms: beyond the causal-historical analysis", *Synthese* 204: 132.

Kitcher, P. (1993). *The Advancement of Science*, New York: Oxford University Press.

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

Kuhn, T. (1970). *The Structure of Scientific Revolutions*, Second Edition, Chicago, IL: The University of Chicago Press.

Ladyman, J. (2011). "Structural realism versus standard scientific realism: the case of phlogiston and dephlogisticated air", *Synthese* 180: 87-101.

Laudan, L. (1981). "A confutation of convergent realism", *Philosophy of Science* 48 (1): 19-49.

Newton, I. (1687). *Philosophiæ Naturalis Principia Mathematica*, the Royal Society in London.

Nola, R. (1980). "Fixing the reference of theoretical terms", *Philosophy of Science* 47 (4): 505-531.

Okasha, S. (2002). "Darwinian metaphysics: species and the question of essentialism", *Synthese* 131: 191-213.

Papineau, D. (1996). "Theory-dependent terms", *Philosophy of Science* 63 (1): 1-20.

Priestley, J. (1775). "An Account of Further Discoveries in Air. By the Rev. Joseph Priestley, LL.D. F. R. S. in Letters to Sir John Pringle, Bart. P. R. S. and the Rev. Dr. Price, F. R. S.", *Philosophical Transactions* 65: 384-394.

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

--- --- (2012). “Causal descriptivism and the reference of theoretical terms”, in A. Raftopoulos and P. Machamer (eds.), *Perception, Realism, and the Problem of Reference*, Cambridge: Cambridge University Press.

Putnam, H. (1975). *Mind, Language and Reality*, Cambridge: Cambridge University Press.

--- --- (1978). *Meaning and the Moral Sciences*, London: Routledge & Kegan Paul Ltd.

--- --- (1981). *Reason, Truth and History*, Cambridge: Cambridge University Press.

Reinikainen, J. T. (2024). “Defending the pure causal-historical theory of reference fixing for natural kind terms”, *Synthese* 203: 131.

Rivat, S. (2025). “How theoretical terms effectively refer”, *Synthese* 205: 154.

Russell, B. (1905). “On denoting”, *Mind* 14 (56): 479-493.

Sanky, C. H. (1994). *The Incommensurability Thesis*, Vermont: Ashgate Publishing Company.

Slater, M. (2015). “Natural kindness”, *British Journal for the Philosophy of Science* 66 (2): 375-411.

Stanford, P. K. and P. Kitcher (2000). “Refining the causal theory of reference for natural kind terms”, *Philosophical Studies* 97: 99-129.

Teller, P. (1995). *An Interpretive Introduction to Quantum Field Theory*, Princeton, NJ: Princeton University Press.

Thomson, J. J. (1897). "Cathode rays", *The Electrician* 39: 104-109.

Wallace, D. (2021). “The quantum theory of fields”, in E. Knox and A. Wilson (eds.), *The Routledge Companion to the Philosophy of Physics*, New York, NY: Routledge.