

Ambiguous Reference

SHAUN NICHOLS
University of Arizona
sbn@email.arizona.edu

N. ÁNGEL PINILLOS
Arizona State University
pinillos@asu.edu

RON MALLON
Washington University-St. Louis
rmallon@artsci.wustl.edu

Abstract:

One of the central debates in the philosophy of language is that between defenders of the causal-historical and descriptivist theories of reference. Most philosophers involved in the debate support one or the other of the theories (or perhaps some combination of them). Building on recent experimental work in semantics, we argue that there is a sense in which both theories are correct. In particular, we defend the view that natural kind terms can sometimes take on a causal-historical reading and at other times take on a descriptivist reading. The meaning will shift depending on the conversational setting. The theoretical view has roots in work by Kitcher (1993). We present some original experiments that support the thesis.

1. Introduction

A theory of reference aims to specify the reference relation that holds between representations and what the representations are about. Identification of the proper reference relation for various classes of terms is the subject of a great deal of work in the philosophy of language, but it is also at the heart of a much wider network of thorny philosophical problems ranging from issues about the nature of mental content (Stampe 1977; Dretske 1981; Fodor 1990) to debates about the possibility of scientific progress (Kuhn 1970; Boyd 1983, 2002), to ontological questions about the existence of entities like race (Appiah 1995; Andreassen 2000), moral facts (Adams 1979; Boyd 1988; 2002), and propositional attitudes (Lycan 1988; Stich 1983, 1996).

Much of this research, both in philosophy of language and in the broader philosophical community, has been guided by classic work on the reference of proper names, theoretical terms, and natural kind terms (Putnam 1973, Kripke 1980; Frege 1893, Russell 1919; Lewis 1972). Two sorts of theories of reference have dominated the stage: descriptivist theories, and causal historical theories.

Descriptivism says that for applicable n , n refers to x whenever (a) competent users of n associate a description D with the term, and (b), x uniquely best satisfies D (Frege 1893, Russell 1919).¹ Later versions of descriptivism require that the referent of a term be the object which uniquely satisfies some relevant cluster or proportion of the descriptions associated with the term (e.g., Searle 1956; Lewis 1972). Descriptivism largely prevailed in philosophy of language until the advent of the causal-historical theory of reference. According to the causal-historical view,

¹ Descriptivism can be conjoined with a theory of meaning, where the description D (or its intensional counterpart) is identified as the meaning (or part of the meaning) of the term n .

for applicable n , n refers to x whenever (a) use of the term n can be traced via a causal-historical chain of communication back to an initial baptism of x (or an otherwise appropriate “grounding”).² Crucially, on the causal historical view, for n to refer to x , competent users need not associate a uniquely identifying description with n . Users could associate no description with the term, or a description that nothing satisfies, or they could associate a description that is not uniquely satisfied, or one that is satisfied by something other than the referent.

In what follows, we characterize issues about reference by contrasting these idealized and influential versions of the descriptivist and causal-historical accounts. These approaches to reference (and a fortiori the specific renderings of them that we offer) certainly do not exhaust the theoretical space.³ Our primary interest is in exploring the nature of folk judgments about the reference of terms and their implications for the theory of reference. We therefore cast our investigation in terms of these prominent models of reference, but with the expectation that the sorts of conclusions we draw will extend to a much broader class of renderings of the reference relation.

² Geach (1969), Donnellan (1970), Kripke (1980), Putnam (1973). Devitt (1981) works out the approach in detail.

³ At the end of section 7 we discuss Gareth Evans’s (1973) hybrid view that incorporates elements from the causal-historical and descriptivist theories. Sam Cumming (2008) and Josh Dever (1991) construe names as similar to variables, and Tyler Burge (1973) treats them as demonstratives. Insofar as these theorists still want to say that names refer, it is not clear that they are either descriptivist or causal-historical.

Because of its centrality, the classic work on reference has been subject to enormous scrutiny, first by two generations of philosophers of language, and more recently, by a growing body of experimental work. How does one go about determining the right theory of reference? Much of the work on theories of reference (both armchair and experimental) proceeds on the assumption that theories of reference must accommodate competent speakers' judgments concerning various classes of terms - proper names, natural kind terms, demonstratives - in various actual and possible circumstances. For example, if a theory of reference entails that in a given scenario, *n* refers to *x*, then the theory is (defeasibly) supported if competent speakers judge that in the scenario *n* does refer to *x* (or something which implies that judgment); the theory is (defeasibly) undermined if competent speakers make the opposite judgment (or something which implies that judgment). Traditionally, theorists (including the authors of classic works on reference in the philosophy of language) have primarily relied on their own judgments, presumably on the assumption that they themselves are representative competent speakers. More recently, experimental philosophers have used experimental methods to measure intuitions about reference, a project that we continue here.

In a series of experiments, we contrast the predictions of descriptivist and causal-historical theories. But our aim is not to vindicate one or the other of these classic views. Rather we offer evidence for the view that both are right – in some contexts, the reference relation for a term *K* is descriptivist and in others, the reference relation for that term is causal-historical. Which reference relation is in effect in a particular situation will depend upon the individual's construal or disambiguation of the term in the situation. This view, which we call an ambiguity view contrasts with both classic views, and also with certain hybrid accounts of reference that

emphasize both descriptive and causal-historical elements in fixing a referent for a class of terms. If we are right, the ambiguity view promises to illuminate longstanding debates within the philosophy of language, and on the much broader range of philosophical debates that rest in part upon assumptions about the correct theory of reference (Mallon et al. 2009).

Here's how we proceed. In section 2, we review some of the extant empirical work exploring folk judgments about reference, and we contrast our own ambiguity view with several alternative views suggested in the experimental and traditional philosophical literatures. Then, in sections 3-6, we present a range of experimental results that support the ambiguity view. In section 7, we consider philosophical precursors for our ambiguity view. Then in section 8, we briefly explore how this view impacts the fate of "arguments from reference" in far-flung corners of philosophy.

2. An ambiguity theory of reference

The experimental philosophical literature exploring ordinary speakers' judgments about reference in various scenarios is now fairly large (e.g., Devitt 2011; Ichikawa et al. forthcoming; Lam 2010; Machery et al. 2009; Marti 2009; Sytsma & Livengood 2011) – too large to review comprehensively here. Instead, we briefly describe three important precedents before setting out our own approach.

An early study by Machery and colleagues (2004) investigated intuitions about scenarios modeled on Kripke's Gödel case. In those cases, the description associated with a proper name turns out to uniquely pick out an individual other than the person originally baptized with the name. In this kind of case, the causal historical theory plausibly entails that the name refers to

the original bearer, and the descriptivist theory entails that the name refers to the person who uniquely satisfies the description. Machery and colleagues found that while U.S. participants tended to give the answer suggested by causal-historical theory, the Chinese participants tended to give the answer suggested by description theory (Machery et al. 2004).

Jylkkä, Railo, and Haukioja (2009) explored intuitions about natural kind terms by presenting participants with vignettes concerning a fictitious natural kind.⁴ They introduced the natural kind term (e.g., ‘zircaum’) and identified it with a certain compound (ACB); they then later indicated that this initial identification was mistaken (the substance is really KML). Although there is a great deal of variation in how participants responded to these questions about natural kind terms, participants favored causal-historical responses over the alternatives (2009, p. 52).

⁴ Jylkka and colleagues’ discussion focuses on externalism and not causal-historicism, but the views are closely related. Externalist theories maintain that the descriptions speakers associate with a natural kind term do not play a necessary role in determining membership in the extension of the term. The extension may instead be determined via some mind-external features of the members of the kind (for instance, that they share some hidden, empirically discoverable essence). This distinction maps onto the causal-historical/descriptivist distinction in familiar ways. The causal-historical theory is a sort of externalist theory or at least as provides a key component for such theories. In contrast, the descriptivist approach is not externalist since according to descriptivism, associated descriptions do play a necessary role in determining the reference of the term.

The third precedent comes from James Genone and Tania Lombrozo. They explored intuitions about reference for natural kind terms by presenting participants with vignettes that varied descriptions and causal-historical relations. Participants were asked to indicate whether two agents were thinking about the same kind (e.g., the same disease). Participants' responses were affected by both descriptive information and causal-historical relations. Genone and Lombrozo argue that these results provide support for a hybrid theory of reference that incorporates both causal-historical and descriptivist considerations within individuals: "A hybrid theory of reference... predicts use of both factors in making reference judgments by the same individual, rather than differential use across individuals" (Genone and Lombrozo, 2012, pp. 722-723).

Our own proposal also maintains that within individuals, both causal-historical and descriptivist factors contribute to a person's intuition about a given case. But our ambiguity theory is not a hybrid view in Genone & Lombrozo's sense. Instead, it is similar to a view that was proposed by Philip Kitcher in his work on theoretical terms in the history of science. Kitcher argues that the language in which scientific theories are presented requires that we adopt a context sensitive theory of reference when we evaluate their claims (Kitcher 1978, p. 524). In some contexts, a token term will refer by virtue of having a "descriptive type" mode of reference; in other contexts, a token of the very same type of term will refer by virtue of having a baptismal type mode of reference (Kitcher 1993, p. 77).

Our view is that the basic pluralist insight here holds not just for the discourse of scientists but also for ordinary people. We will argue that natural kind terms are ambiguous. In

some cases, the reference of a token is fixed by a causal-historical convention⁵; in other cases, the reference of a token of the same type is fixed by a descriptivist convention.⁶ We call this an ambiguity theory because the idea is that there are two conventions that determine the reference of natural kind terms, much as there are two conventions that determine the reference of the word ‘bank’, though, as we will see, there are crucial differences between the ambiguity found in ‘bank’ and the one we think is operative in natural kind terms”.⁷

The method we use to make our case is traditional in the sense that we look to judgments employing natural kind terms in various scenarios. As we noted at the outset, such reactions are

⁵ We adopt the expression “reference convention” to supply us with a term that does not presuppose that the correct account of reference will posit a univocal reference relation. But we do not mean to tie our view to any particular theory of convention. We might also think of reference relations as ‘practices’, ‘modes’, or ‘construals’.

⁶ Our ambiguity theory also differs from “dual aspect” semantics theories which posit a two tiered theory, where one tier concerns understanding and the other reference. Our theory posits ambiguity in the referential aspect.

⁷ This ambiguity differs from the kind of context sensitivity exemplified by ‘tall’ or ‘I’. For those kinds of terms there is a single convention which determines how reference gets fixed depending on the context. Note that it’s possible that in addition to the ambiguity of natural kind terms, they can also be context sensitive (in the sense of ‘tall’ and ‘I’). For instance, on the descriptivist convention of reference fixing, there might be more or less exacting standards for satisfaction. In some contexts, perhaps all of an associated description must be met; in other contexts, perhaps only a small portion of the description must be met.

commonly used as evidence for the truth of one or another substantial account of reference, and we join in the recent trend of using experimental methods to get at these judgments. However, we developed our experiments to address some of the shortcomings of the work reviewed above.

One issue that has hounded many of the experiments exploring reference is the possibility that the experimental materials confound semantic reference with something else – most often speaker’s reference. For instance, in Machery et al, participants are asked to indicate whom John is talking about when using the name ‘Gödel’. This has led a number of commentators to argue that when people give apparently descriptivist responses, this is because they are really answering the question “What is the speaker trying to refer to?” as opposed to the properly semantic reference question, “What does the name refer to?” A similar complaint might be directed at Genone & Lombrozo, who ask what the agents are thinking about. It is, of course, a substantive empirical claim that those who give apparently descriptivist responses are answering the speaker’s reference question, and there is a lively debate about who is right here (e.g., Deutsch 2009; Ludwig 2007; Sytsma and Livengood 2011; Ichikawa et al. 2012; Machery et al. forthcoming a). But it is a *prima facie* limitation of the experiments if they invite judgments concerning speaker’s reference.

A more general problem afflicting all of the extant work is that the experimental vignettes are quite complex (cf. Devitt 2010, p. 419). They involve elaborate hypothetical scenarios with competing descriptions and multiple agents. In Machery et al. (2004), participants are told about a guy (John) who learns about another guy (Gödel) who proved some theorem; then they’re told to imagine that unbeknownst to John this “Gödel” didn’t really prove that theorem, but some third guy (“Schmidt”) did; then they’re asked about John’s use of the name “Gödel”. In Jyllka et

al. (2009), participants are told about some mineral called “zircaum” that scientists think is the compound ACB; then scientists find a mineral someplace else that has the “deep structure” ACB; but wait! - the first stuff turns out not to be ACB. It’s really this other deep structure - KLM. But there’s more - the stuff found in the second place really is ACB. Genone & Lombrozo’s task is no simpler – the questionnaires themselves are longer than many articles in *Analysis*. We ourselves have been involved in some of this work, and so we are well aware of why the materials get so complex. It is very hard to get materials that are (i) targeted to the philosophical issue, (ii) comprehensible, and (iii) concise. The road is difficult, and we think the work has been illuminating. But there can be no doubt that it would be better to have simpler materials to get at the questions of interest.

In order to skirt concerns that previous materials indicate speaker’s reference rather than semantic reference, in these materials we do not even ask about people’s thoughts. We focus instead on direct questions using a kind term. To address the complexity of vignettes and eliminate the purely hypothetical nature of the tasks, we turned to actual cases of contentious kind terms from natural history. As it happens, Medieval bestiaries provide a wonderful resource for contentious species kinds (see, e.g. Bartholomaeus Anglicus 1482; Isidore of Seville 2006; Pliny the Elder 1940). Such bestiaries often have accurate descriptions of species, but they also often grossly misdescribe known species and sometimes describe animals with truly bizarre traits. This afforded us the opportunity to present participants with real examples of kind terms drawn from Medieval natural history. And rather than invoke multiple agents (and their thoughts), we ask straightforward, non-hypothetical, questions concerning reference, existence, or predication. As a result, the materials are much simpler than those used in previous studies.

In the current project, we want to offer evidence for the ambiguity theory by exploring whether different manipulations will affect people's reference judgments about kind terms. Broadly speaking, our strategy for developing different manipulations relied on the kinds of philosophical arguments used to defend descriptivist and causal-historical theories of reference. The idea is that in defending one or the other theory of reference, philosophers deploy presentations and examples that guide our intuitions to favor the author's preferred theory-- which on our view turns out to be just a disambiguation of the term at issue as opposed to an argument that settles the matter for all uses of the term. For instance, famously, causal-historical theories do better at accommodating the presumed continuity of reference in the history of science. By contrast, descriptive theories do better at accommodating failed reference at least in certain domains.⁸ Thus, in testing the ambiguity theory, we constructed cases that exploited these kinds of theoretical advantages to see whether people's intuitions would be affected in predictable ways.

3. Study 1: Mistakes were made

For our first task, we drew on the fact that causal-historical reference provides a natural way to explain continuity of reference through (sometimes profound) theory change in the history of science (Boyd 1983). The basic idea is that often in the history of science, as we come to discover that earlier views implicating a kind are mistaken, we retain the kind term and revise the description. It is commonly taken to be an advantage of causal-historical theories that they can accommodate this practice (since they do not presume the description – now falsified – plays a

⁸ See, e.g., Devitt & Sterelny 1999; Lycan 2008.

role in determining the referent of the term). Thus, in devising our first experiment, we predicted that descriptivist intuitions would be deflated if we exposed participants to a case in the history of science in which scientists mischaracterized a kind, even though we in fact presume a continuity in reference. The idea is that such examples naturally induce a causal-historical way of thinking about reference, and our prediction is that by bringing such an interpretation to mind, this will make participants less inclined to give a descriptivist interpretation to a subsequent example. We conducted a study to test this prediction.

When Kripke and Putnam push the idea that reference is preserved across mischaracterization, they tend to use fictional examples – we are to imagine various mistakes in the identification of gold or tigers or cats. But for our experiment we wanted to use a real example from the history of science. Triceratops provides an excellent case. As it happens, Triceratops was initially characterized in ways that we now think provide a dramatic mismatch. In one condition, participants were presented with a description that did not emphasize these mistakes:

Neutral condition:

The dinosaur Triceratops has a giant bony structure, a “frill”, behind the head. As scientists have accumulated more fossils, they have learned much more about this dinosaur. For instance, as the Triceratops ages, the frill becomes much longer. The frill also develops giant holes and these holes become covered with keratin, which is a key component in human skin. Researchers have argued about different ways that the frill might become longer and develop holes. This remains an issue of discussion in dinosaur research.

The other condition noted that the Triceratops was mischaracterized in a dramatic way, while implying continuity of reference:

Mischaracterization condition:

The Triceratops is a large dinosaur with a giant bony structure, a “frill”, behind the head. However, our understanding of the Triceratops has changed dramatically over the last century. As scientists have accumulated more fossils, they have come to recognize that there were many mistakes made in the initial description of this dinosaur. For instance, scientists thought that the Triceratops had skin like an elephant, but it turns out that it really had scales like an alligator. They thought that Triceratops was exclusively a plant eater, but now hold that the Triceratops was at least partly a meat-eater. And most strikingly when scientists first named Triceratops, they thought it was an ancient bison. Only later did they realize that it was really a dinosaur.

In both conditions, participants were then presented with a completely different case. The case was drawn from medieval bestiaries. The case involved a putative animal called ‘catoblepas’ (Pliny the Elder 1940, book 8), and all participants read the following description:

In the Middle Ages, animal researchers described a distinctive kind of mammal. They called it catoblepas. The catoblepas was said to be like a bull but with a head so heavy that the animal has to keep its head down at all times. It was also thought that the catoblepas had scales on its back. In addition, the researchers said that looking into the animal’s eyes causes immediate death. Of course there is nothing that meets this description, but researchers think that it was based on reports of encounters with wildebeest.

Following this description, participants were asked whether catoblepas existed. We phrased the existence question in terms of clear cases:

Obviously there are some species that really exist and some that really don't. Rabbits really exist; goblins don't really exist. Please indicate to what extent you think catoblepas are more like rabbits or goblins in this regard.

Are catoblepas more like rabbits (really exist) or goblins (don't really exist)?

Participants responded on a six-point scale, with endpoints marked by goblin on the left and rabbit on the right. Participants in the neutral condition tended to give descriptivist answers – saying that catoblepas are more like goblins (don't really exist) ($M=1.94$ out of 6). The average response in the mischaracterization condition ($M=2.97$) was significantly less descriptivist ($N=63$; $t(49) = -2.772$, $p < .01$).

This first result fits naturally with the ambiguity theory. By providing a frame that emphasizes continuity of reference in the history of science, we deflate descriptivist responses. But this study only looks at one factor, and to support the ambiguity theory, we wanted to look at a broader range of factors.

4. Study 2: Existence vs. Reference

Perhaps one of the most important objections to causal-historical theories concerns empty terms. Sometimes, the causal-historical path from a modern use of a term may lead back to an object even though, intuitively, the term has no referent (or refers to a different object). For example, the modern use of 'Santa Claus' in the United States may trace back to the historical figure Saint Nicholas (4th century). Yet, we don't think that Santa Claus exists or that 'Santa Claus' refers to

the historical figure.⁹ Presumably, part of the explanation for this is that nothing comes close to satisfying the description associated with the term. Following this line of thought, we predicted that the interpretation of existential statements involving terms whose associated descriptions are not satisfied by anything, will tend to be disambiguated using a descriptive theory of reference. The following study attempts to confirm this idea.

In the neutral (control) condition of Study 1 we saw that subjects tended to say that catoblepas (a medieval term causal-historically linked to wildebeests) don't exist. This fits with the idea we just discussed that the existence claim invites a descriptivist interpretation. But what would happen if we asked people whether 'catoblepas' refers to wildebeests? Our hypothesis is that this statement will be more likely to elicit causal-historical intuitions, since the statement makes salient an obvious candidate for the causal-historical referent of 'catoblepas', viz., wildebeests. We tested this comparison for this second, exploratory study. We used the catoblepas example again, this time with the most unusual element from the description (i.e., the death gaze) eliminated. Participants read the following:

In the Middle Ages, animal researchers described a distinctive kind of mammal. They called it catoblepas. The catoblepas was said to be like a bull but with a head so heavy that the animal has to keep its head down at all times. It was also thought that the catoblepas had scales on its back. Of course there is nothing that meets this description, but researchers think that it was based on reports of encounters with wildebeests.

⁹ The cultural context matters here. In Germany, for instance, the term "Saint Nicholas" might refer to the historical figure. We thank the editor for this observation.

Following the description, participants were asked to indicate their level of agreement (1=strongly disagree; 6= strongly agree) to two statements, counterbalanced for order:

1. “Catoblepas” refers to wildebeests.
2. Catoblepas exist.

The existence question tended to produce responses ($M=3.06$) which were more descriptivist than the responses to the “reference” question ($M=3.68$). The difference was statistically significant ($t(33) = 3.12, p < .01$).¹⁰ That is, people are more inclined to say that ‘catoblepas’ refer to wildebeest than that catoblepas exist. These results are puzzling since if ‘catoblepas’ refers to wildebeests, then catoblepas exist. So people should not record a higher level of agreement with the former than with the latter. The ambiguity theory can resolve the puzzle: The “refers to” question, with the explicit mention of wildebeests, invites a causal-historical interpretation. And the “exist” question invites a descriptivist interpretation of the kind term (see also Machery et al. forthcoming b). But this is just a preliminary result. We investigate this further in Study 3.

5. Study 3: Existence, Presupposition and Predication

In previous studies [Study 1 (Neutral condition) and Study 2] we saw that people tended to disagree with the statement ‘Catoblepas exists’. This is difficult for the causal-historical theory to explain since the theory seems to predict that ‘catoblepas’ refers to wildebeests. In study 3a we

¹⁰ A few more statistical details: There were 34 participants (after excluding 6 participants who didn’t know that the wildebeest is a real species and 7 who had taken a similar survey). There were no effects of the order in which the questions were asked. (Reference question: $t(32) = -.37, p = .739, n.s.$; Existence question: $t(32) = -.24, p = .8, n.s.$).

explore this idea further. Using the same vignette as in study 1, we ran a between subjects experiment in which we asked participants to indicate their level of agreement with either (1) a statement that predicated “are wildebeests” on “Catoblepas”; (2) a statement that predicated “are catoblepas” on “Wildebeests”; or (3) an existence statement:

(1) Catoblepas are wildebeests

(2) Wildebeests are catoblepas

(3) Catoblepas exist.

According to the causal-historical theory, ‘catoblepas’ and ‘wildebeests’ are presumably coreferential, so defenders of that view should expect responses to (1) and (2) to be the same. We did not find this. Responses differed. Participants tended not to disagree with (1), that catoblepas are wildebeests ($M=3.67$ on a 6 point scale, 3.5 is the midpoint). But they disagreed with (2), Wildebeests are catoblepas ($M=2.79$). The difference was statistically significant ($t(41)=2.39$, $p<.05$).¹¹

From the perspective of a univocal approach to reference, the data is puzzling no matter what one thinks of the causal-historical theory. There are two related mysteries. First, it is puzzling why responses to (1) and (2) diverged.¹² Second, it is mysterious that responses to (3),

¹¹ The total number of participants was 60 (after excluding 10 who didn’t know that the wildebeest is a real species and 4 who had taken a similar survey). There was no difference in responses between (2) and (3), $t(34)=.06$, $p=.956$.

¹² One interpretation of this pattern is that participants think that catoblepas are a proper subset of wildebeest. Our current data do not rule this out and we acknowledge that more research is needed here.

which land on the disagreement side ($M=2.76$) differ significantly from responses to (1) which do not land on the disagreement side ($t(39)=2.15, p<.05$). How can people disagree with (3) and not with (1)? One would have expected (1) to entail (3).

And the mystery deepens. In a follow up study, (3b), we used a within-subjects design, giving each participant questions (1) and (3) one right after the other (counterbalanced for order). If (1) obviously entails (3), then subjects should see this when the statements are juxtaposed. Their responses should thereby reflect this. But again, we did not discover this. Responses to (1) and (3) were significantly different $t(33)=2.633, p>.05$. Responses to (1) again crossed the threshold towards agreement ($M=3.53, SD=1.461$) and responses to (3) were again on the disagreement side ($M=2.88, SD=1.629$). So even when the cases are juxtaposed, it does not seem like subjects are responding as if it is flatly inconsistent to affirm (1) and deny (3).¹³

The ambiguity hypothesis explains the results by proposing that the referential mechanism for ‘catoblepas’ in (1) tends to be causal-historical while the referential mechanism for that word in (2) and (3) tends to be descriptivist. The ambiguity of ‘catoblepas’ explains why subjects do not exclude the possibility that (1) is true while (3) is false.

Obviously it is the barest sketch of a theory to say that the interpretation of ‘catoblepas’ in (1) is causal-historical whereas the interpretation in (2) and (3) is descriptivist. But we think a little more can be said. It’s plausible for independent reasons that the fact that ‘catoblepas’ occurs in the subject position of (1) triggers a presupposition that catoblepas exist (see, e.g.,

¹³ There were no detectable ordering effects. There was no difference between responses to statement (1) across the orderings, $t(32)=.34, p=.75$ n.s., and no difference between responses to statement (3) across the orderings, $t(32)=.44, p=.66$ n.s.

Strawson 1964; Reinhart 1981; Lasersohn 1993, von Fintel 2004).¹⁴ Our subjects accommodate this presupposition. And this explains the general failure to disagree with (1). Since (1) presupposes that catoblepas exist, our subjects seek to find an available interpretation of ‘catoblepas’ in which the term has a referent. The descriptivist interpretation for reference fixing is unavailable, since on that interpretation, the term has no referent. Subjects are instead moved to exploit the causal-historical interpretation for reference fixing, since on that interpretation ‘catoblepas’ plausibly refers to wildebeests. This makes accommodation possible. In this way, we can explain the lack of disagreement with (1).

In sharp contrast, (2) and (3) do not presuppose that catoblepas exist. In the case of (3), it is not presupposed that catoblepas exist simply because this is exactly what the claim says (so it cannot also be what is presupposed). In the case of (2), it is not presupposed that catoblepas exist because ‘catoblepas’ occurs in the predicate position and there is independent reason to think that

¹⁴ That the presupposition is triggered in this context is a natural position to take. Both “Cateblopas are wildebeests” and the negated “Cateblopas are not wildebeests” commit the speaker to the existence of cateblopas, in ordinary contexts. Similarly, asking “Are cateblopas wildebeests?”, ordinarily assumes that cateblopas exist. All of this is evidence for the presupposition claim. It is difficult to say exactly why ‘cateblopas’ in the subject position of (1) triggers a presupposition of existence if it does. Strawson (1964) and Reinhart (1981) would trace this to the topic-comment features of the utterance. Other researchers such as Lasersohn (1993), and von Fintel (2004) focus on verifiability and background information. We are not committing here to an explanation of why the presupposition is triggered. We are only claiming that it is in fact triggered in (1).

predicates – unlike subjects – do not carry a presupposition of existence.¹⁵ We propose instead that ‘catoblepas’ in (2) and (3) tends to be associated with the descriptivist mechanism for reference fixing. This explains why subjects in fact tend to disagree with those statements. Nothing satisfies the description associated with that term.¹⁶

Thus, the results of this study put pressure on a univocal theory of reference along the lines of the causal-historical theory. But the results can easily be explained by appealing to an ambiguity theory, on which terms may sometimes take on a descriptive interpretation and at other times, a causal-historical interpretation.

¹⁵ Strawson (1950) notes that when “The greatest French soldier” appears in predicate position in a sentence like “Napoleon is the greatest French soldier”, the expression is not used to mention any particular individual and hence presumably does not trigger a presupposition of existence. By contrast, if the description appears in the subject position in a sentence such as “The greatest French soldier died in exile”, it is used to refer to a particular individual and it triggers a presupposition of existence.

¹⁶ Why couldn’t the presupposition, rather than the causal-historical use of natural kind terms, give rise to agreement with “Catoblepas are wildebeests” (and hence explain the data)? Not all presuppositions can be accommodated. If I say ‘the king of France is coming’ you will reject this, not accommodate it. So it is not enough to explain agreement with (1) to say that it presupposes existence. There must also be available a reading of ‘catoblepas’ which can accommodate it. The descriptivist reading won’t do this, but the causal historical theory does.

6. Objections: Study 4

We would like to discuss two objections to these studies. First, it might be thought that the vignettes are so underdetermined that the causal-historical views do not make a clear prediction about terms like ‘catoblepas’. Hence, responses to the surveys cannot count as evidence for or against the causal-historical theory. The objection says that if only the causal-historical element of the vignette were made salient, responses would fall in the causal-historical side of the scale.¹⁷ Second, and relatedly, it might be thought that in order to show that natural kind terms are ambiguous we have to show that responses shift among frames between clear agreement with the causal-historical response and clear agreement with the descriptive response.¹⁸ In our probes, we show significant shifts in the predicted direction but we do not often see clear agreement with the causal-historical response. Instead, we often see shifts between clear agreement with the descriptivist response and a neutral response in between descriptivist and causal-historical perspectives.

To deal with these objections we conducted one last study. We modified the catoblepas probes to make the causal-historical elements more salient – the new vignette says that many scientists, when faced with wildebeests, would often call them ‘catoblepas’. Here’s the full vignette:

¹⁷ Thanks to Michael Devitt and Genoveva Marti on this issue.

¹⁸ Note, for instance, that despite significant differences between conditions in our earlier studies, the “causal-historical” responses often do not differ significantly from midline. We would like to thank Edouard Machery for pressing us on this point.

In the Middle Ages, animal researchers described a distinctive kind of mammal. They called it catoblepas. The catoblepas was said to be like a bull but with a head so heavy that the animal has to keep its head down at all times. It was also thought that the catoblepas had scales on its back. In addition, the researchers said that looking into the animal's eyes causes immediate death. Of course there is nothing that meets this description, but researchers know that it was based on reports of encounters with wildebeest. Many scientists in the middle ages, when faced with wildebeest, would often call them catoblepas.¹⁹

In a within-subjects study, participants were asked the extent to which they agreed with “Catoblepas exist” and also “Catoblepas are wildebeest”.²⁰ Options ranged from (1-6) where (1) is “Strongly Disagree” and (6) is “Strongly Agree”, as before. Consistent with our other probes, responses to the existence probe were still descriptivist (despite the modified probe to presumably favor causal-historical interpretations) (N=32, M=2.88, SD=1.6). The mean was significantly below the neutral point ($t(31)= 2.208, p=.035$). On the “Catoblepas are wildebeests” statement, participants showed clear agreement (N=32, M=4.25, SD=1.6). The mean was

¹⁹ We recognize this vignette risks confusing use and mention since ‘catoblepas’ does not appear in quotation marks. We take it this risk is not pernicious since lay subjects do not strictly observe this convention, and since there is no obvious interpretation on which the word is used.

²⁰ Questions were counter-balanced. There were no ordering effects. Subjects who failed a comprehension check were excluded from the statistical analysis.

significantly above the neutral midpoint ($t(31)=2.641, p=.013$). This indicates that for this second probe, answers fell clearly in the causal-historical side of the scale.²¹

These results help us to meet the objections. The first objection is met because we still managed to get descriptive responses despite having made more salient the causal-historical elements in the vignette. The second objection is met because with this new probe we are able to get people to shift from clearly descriptive to clearly causal-historical responses.

7. Ambiguity, Inconsistency, and Hybrids

Our experiments show a range of contextual effects on judgments that implicate reference. We have argued that these results support an ambiguity theory of reference. Contextual effects are sometimes interpreted as showing rational failings (e.g., Tversky & Kahneman 1981). If participants agree that “Catoblepas are wildebeests” but deny that “Catoblepas exist”, then that sure looks like an inconsistency. But we think that the ambiguity theory provides a better account of the phenomenon. One reason to think this is that subjects tend to give these responses even when the questions are placed back to back (suggesting that they don’t sense these responses as inconsistent).²² Rather than conclude that people make systematically inconsistent judgments about natural kind terms, we think a more plausible and charitable view is that natural kind terms are ambiguous.

²¹ A paired sample T-test reveals that the responses to the probes are significantly different ($t(31)=5.1222, p<.01$).

²² By contrast, in the flagship example of rational inconsistency - the Asian Disease case - placing the corresponding questions back to back (within subjects) significantly reduces the rate of “inconsistent” responses. See Stanovich 1999, p. 107.

Consider the most pedestrian case of ambiguity. It is not inconsistent to agree with “Banks hold money” as uttered in some contexts and disagree with it as uttered in other contexts. For the context can serve to disambiguate the meaning of the word ‘bank’ as financial institution in one context and riverside in another. Similarly, in the present case, context disambiguates whether a kind term refers via its associated description or by its causal-historical connections.²³

Again, our proposal is that natural kind terms (and plausibly names as well) are ambiguous, such that in some cases the reference is determined descriptively and in other cases non-descriptively. This suggestion runs against just about everything in the literature. In the philosophy of language, work on the theory of reference has operated under an assumption that

²³ Our use of ‘context’ here must be understood in a permissive sense which will include information about the speaker’s intentions. It is not to be understood in a more restrictive sense which would make ‘bank’ context sensitive in the sense that the word ‘I’ is context sensitive. We do not think, for instance, that the word ‘bank’ has a Kaplanian character which is a function from contexts to intensions.

²⁴ Many philosophers of language do maintain that different theories of reference apply to different classes of terms (e.g. terms for artifacts and terms for natural kinds), but few maintain that different theories of reference apply to the very same kind of term in different contexts.

only one theory of reference will apply to a class of terms.²⁴ This assumption has been a critical constraint on theory building in philosophy of language.²⁵

Our ambiguity theory thus abandons a dominant line of thought across a wide swath of philosophy. Indeed, in a related context, Kripke offers a cautionary note: “It is very much the lazy man’s approach in philosophy to posit ambiguities when in trouble” (Kripke 1977, p. 268).

We take comfort that in our indolence, we have a venerable co-traveler in the philosophy of science – Philip Kitcher. In keeping with Kitcher, we resist dismissing the folk responses in our experiments as a kind of silly inconsistency. The folk responses, we believe, suggest that natural kind terms are ambiguous.

So what determines whether a speaker, on a particular occasion, will felicitously use either the descriptive or causal-historical disambiguation of a natural kind term? We are not optimistic that there will be a well-defined solution to this problem. But notice there is no well-defined solution for how speakers resolve canonical cases of ambiguity like ‘bank’ either. The best answer to this question may just be that, in general, the word ‘bank’ (on a particular

²⁵ In an earlier paper (“Against arguments from reference”) two of us (Mallon & Nichols) reject the assumption that there is a single set of reference intuitions in the population. In that paper, we allowed that different people might have intuitions that support different theories of reference (Mallon et al. 2009, p. 344). But we did not explore the possibility that within each of us, there are (at least) two ways of thinking about the reference of kind terms.

occasion) will felicitously refer to whatever the speaker chooses to talk about given his conversational goals.²⁶ However, as we explain below, in many typical uses of natural kind terms it will simply be indeterminate which reference fixing mechanism is in play. Even if there is no well-defined solution to disambiguation, we might still want to seek out features in the conversation that lead an agent to invoke a particular disambiguation of a term. For certain cases, if an agent uses a term in what is obviously the causal-historical sense, other conversational participants will follow or “accommodate” her (cf. Lewis 1994, p. 424). If the agent uses it in the descriptive sense, the participants will follow her there as well. There are limits, of course, as participants won’t follow a speaker no matter what.

Turning back to our experiments, our explanation for why certain prompts got a causal-historical reading can now be re-described as involving accommodation. For instance, we saw that people tended to agree more with “Catoblepas are wildebeests” than with “Catoblepas exist”. Our explanation was that the existence of catoblepas is presupposed in the former but not the latter. Given the information in the vignette (where nothing fits the description associated with ‘catoblepas’), the presupposition cannot be accommodated by assuming a descriptivist use of ‘catoblepas’, hence we think that subjects accommodate by interpreting the term in a causal-historical sense. In contrast, the second sentence does not carry that presupposition, so there is less pressure to give ‘catoblepas’ there a causal-historical interpretation, and so there is less chance people will agree with that statement (which is the result that we got).

²⁶ We say “in general” because even if a speaker wants to use ‘bank’ to mean riverside, if the conversation he is engaged in is decidedly about financial institutions, his use of ‘bank’ may not felicitously refer to the riverside.

Not all factors that cause shifts in referential interpretations can be explained by appealing to this kind of accommodation. Consider, for instance, our first study. It is well known that descriptivism has difficulty explaining scientific progress. Scientific progress would seem to require continuity of reference even in cases where beliefs about the purported referents of the theoretical terms have undergone radical change involving progress. In study 1, we call attention to such cases by describing the mischaracterization of Triceratops. And when the folk are primed with such cases, they subsequently lean towards non-descriptivist responses. Our interpretation of this is that by priming people to think in a causal-historical-way, we “bias” their subsequent interpretation of natural kind terms. If that is the right interpretation, it suggests that the features of a conversation that push speakers to disambiguate a natural kind term in a certain direction may involve a lot more than accommodation.²⁷

Having said this, we do not mean to suggest that the ambiguity at issue is just like the straight-forward lexical ambiguity of ‘bank’. The meanings of ‘bank’ are semantically unrelated. It is no surprise that in other languages these meanings are expressed with distinct word forms. In sharp contrast, the causal-historical use of ‘tiger’, say, is semantically related in an interesting way to the descriptive use of that term. We would not expect other languages to express these

²⁷ For this reason, we prefer to describe the observed patterns of dual use as cases of ambiguity, as opposed to trying to account for them as an outcome of a unified hybrid concept. A unified hybrid account would explain the shifts in uses of ‘catoblepas’, for instance, as following from the meaning of that term. We don’t readily see what such a meaning for ‘catoblepas’ would be. By contrast, our account is relatively simple. We account for the shifts in usage by appealing to our ordinary capacity to disambiguate.

meanings by utilizing distinct word forms. It's worth noting that Lewis' account of vagueness as semantic indecision also works like this (Lewis 1999). For instance, Lewis argues that the expression 'house' is ambiguous between a main structure of a house-like building and a main structure plus a garage.²⁸ Clearly, the ambiguity here is unlike the ambiguity of 'bank' in that the candidate meanings are very much related. We think the proper model for the ambiguity of natural kind terms is much more like Lewis's "semantic indecision" cases than the simple cases of ambiguity exemplified by 'bank'. As a consequence, we do not expect that standard linguistic tests that detect lexical ambiguities (like that involving 'bank') to apply to natural kind terms any more than they would apply to Lewis' 'house'.

If we follow Lewis' model for semantic indecision, an interesting feature of our approach falls out. In contrast with typical uses of 'bank' where a speaker usually has one of two readings in mind, when people are using uncontested natural kind terms, it's far from clear that they intend one of the reference conventions rather than the other. In typical uncontested cases of natural kind terms, the causal-historical and the descriptive mechanisms are in "harmony".²⁹ When I say, "There is water in the Hoover dam," what I say is true under both reference

²⁸ Lewis 1999, p. 172.

²⁹ We thank the editor and a referee for emphasizing the significance of these cases, as well as for the term "harmony".

conventions for “water”.³⁰ So in this case, the lack of a determinate answer to “which reference convention is operative” is harmless. Lewis’ model of semantic indecision provides a natural way to accommodate this.³¹

We now turn to address whether a “hybrid theory” like Gareth Evans’s can explain our data. According to Evans, the denotation of a name will be the dominant source of the information (descriptions) associated with the name. The view is partly causal-historical because

³⁰ There is a subtle issue about the nature of the harmony in uncontested cases. Harmony can happen in a couple of ways.

- i. One possibility is that in harmonious cases, the two ways of fixing reference (causal-historical and descriptive) both lead to the exact same extension.
- ii. Another possibility is that in harmonious cases, the extensions overlap in the cases that matter to us, without being exactly the same extensions. So, a causal-historical treatment would maintain that “water” refers to H₂O and a descriptivist treatment might maintain that “water” refers to “drinkable liquid in rivers etc”. These assignments don’t have the same extension because in Twin earth (assuming it is real) there is no H₂O. But it doesn’t matter since we never talk about Twin earth in the real world.

On either of these possibilities, normal sentences involving “water” would be true on either interpretation for most uses that matter. This is similar to “house” where if I say “that’s a beautiful house”, this would be true whether the interpretation means “main structure” or “main structure with garage”.

³¹ Those seeking a formal treatment can make use of a supervaluational semantics.

whether an object counts as a source of information will be a causal matter and not a matter of satisfying some condition included in the information associated with the name. The view is partly descriptivist because reference is mediated by the description associated with a name.

Evans' theory arguably gets it right in some cases where Kripke's gets it wrong. Evans considers a case where ancient documents are discovered containing interesting mathematical proofs. Inscribed in these documents is the name 'Ibn Kahn' who is now mistakenly taken to be the author of the proofs. In fact, the person originally named 'Ibn Kahn' was the scribe. Evans intuitively thinks that present uses of 'Ibn Kahn' refer to the author of the proof, not the scribe, contrary to what Kripke would presumably predict. According to Evans, the reason for this is that the dominant source of the information associated with 'Ibn Kahn' would be the mathematician not the scribe. This is because the information we associate now with the name mostly concerns mathematical proofs which is better seen to be causally connected to the author as opposed to the scribe.

Can this hybrid account explain our data? What is crucial to our data is that people can readily switch back and forth between two interpretations of natural kind terms. As the first experiment shows, this can happen simply by getting them to think about the possibility of mistakes. As the subsequent experiments show, this can happen simply by switching the question being asked of them (Instead of asking them about whether Catoblepas exist, we ask them about whether Catoblepas are wildebeests). The hybrid account can explain these data only if the subtle changes in conversational context are enough to change what counts as the dominant source of the body of information associated with the name. It's not at all clear that Evans' theory entails that these subtle changes are sufficient. In our experiments the descriptions associated with the

name ‘Catoblepas’ are always the same. But how can it be that in some conditions *wildebeests* are the dominant source of information and in the other *nothing* is the dominant source of that same information? This would happen only if somehow what counts as a dominant source of the information varies with subtle contextual changes.

We are open to the possibility that what information one (readily) associates with a name can vary from context to context. And perhaps in those situations, Evans could predict variance in referential intuitions across subtle changes. But crucially, this is not how our cases are set up. The advantage of using the Catoblepas case is that users do not come with prior understanding of the term. And hence the only description they associate with the name is the one that we give them, which (as we mentioned above) is constant across the conditions.

In contrast, Evans’ own examples of referential shifts concern different populations associating different descriptions with the term at issue. For instance, when modern people use ‘Ibn Kahn’, they refer to the mathematician, not the scribe. But when the scribe’s contemporaries used the name, they referred to the scribe. Obviously, these populations associated radically different descriptions with the names. Again, this is not how our experiments are set up. For us, the description associated with ‘Catoblepas’ remains fixed.

In addition, Evans explicitly rejects the idea that denotations of names can shift from moment to moment depending on subject matter:

I think we can say that *in general* a speaker intends to refer to the item that is the dominant source of his associated body of information. It is important to see that this will not change from occasion to occasion, depending on subject matter (Evans 1973, p. 17)

Immediately following this, he clarifies what he means by “subject matter” by using an example where what is predicated of a name changes. So that if in one breath we say “Napoleon is x” and in the next breath, we say “Napoleon is y” (where the subject matter of x is distinct from y), then there won’t be a shift in reference. But this view conflicts with our within subjects results (experiment 3b) where we seem to get a reference shift across “Catoblepas exists” and “Catoblepas are wildebeests”.

So, although there are important elements of Evans’ view that inspire our own view, we think that Evans’ view does not predict the results of our experiments, and sometimes predicts different results. Of course, there may be further elaborations of his view that predict our results. In addition, we are open to the idea that the ambiguity, when fully worked out, may involve elements of a hybrid view.

8. Ontology

The previous section focused on how our results might have implications for philosophical theories of reference. In particular, we argued that our data support a theory on which kind terms are ambiguous between descriptivist and causal-historical forms of reference.

As noted in the introduction, reference also plays a critical role in arguments across a range of other philosophical subfields. Since a large measure of the motivation for this project was the role of reference in ontological disputes, we want to consider, too briefly, how accepting referential ambiguity impacts these disputes. Ontological disputes arise concerning everyday

terms like “race” and “belief”, as well as theoretical terms like “phlogiston” and “black hole”.³²

Consider belief. Eliminativists argue that beliefs do not exist, and we can characterize their argument as follows (Mallon et al. 2009, p. 333):

1. propositional attitude terms like “belief” and “desire” refer via the implicit description associated with them by our folk theory of mind.
2. there is nothing that has all (or most, or the most important) of the properties this description associates with “belief” and “desire.”
3. beliefs and desires don’t exist.³³

In an insightful discussion, Stephen Stich (1996) draws attention to the role of assumptions about reference in eliminativist arguments in the philosophy of mind (Churchland 1981; Stich 1983), noting that the eliminativist argument assumes some descriptive account of the reference of

³² As we will see below in the case of “belief”, some philosophers maintain that the term refers to a natural kind; others maintain that the term doesn’t refer to anything. Note that it’s consistent to maintain that a term is a natural kind term while also maintaining that the term fails to refer to a natural kind (see, e.g., Leslie forthcoming).

³³ As Bishop and Stich (1998) point out, this argument requires another premise, one that licenses the inference from “‘belief’ does not refer” to “belief does not exist”. We set this issue aside here.

propositional attitude terms.³⁴ We focus on the case of belief, but the type of eliminativist argument Stich draws attention to can be found in a range of domains. (See, for example, the assumption of descriptivism in the eliminativist arguments against the existence of race [Zack 1993, Appiah 1995] or against objective values [Mackie's 1977] or against beliefs [Stich 1983].)

Although this line of argument is familiar, on alternative assumptions about the correct reference relation, the conclusion does not follow. For instance, Lycan rejects the eliminativist arguments about belief precisely because they presuppose a descriptivist theory of reference, writing:

I am at pains to advocate a very liberal view... I am entirely willing to give up fairly large chunks of our commonsensical or platitudinous theory of belief or of desire (or of almost anything else) and decide that we were just wrong about a lot of things, without drawing the inference that we are no longer talking about belief or desire. To put the matter crudely, I incline away from Lewis's Carnapian and/or Rylean cluster theory of the reference of theoretical terms, and toward Putnam's ... causal-historical theory. As in Putnam's examples of 'water,' 'tiger,' and so on, I think the ordinary word 'belief' (qua theoretical term of folk psychology) points dimly toward a natural kind that we have not fully grasped and that only mature psychology will reveal. I expect that 'belief' will turn out to refer to some kind of information-bearing inner state of a sentient being, ... but the

³⁴ There are, of course, ways to challenge this argument that have nothing to do with reference. For instance, one might maintain, with Fodor (1987), that the description of belief in the folk theory is largely right; or one might maintain, with simulation theorists, that there is no folk theory of belief (Gordon 1986).

kind of state it refers to may have only a few of the properties usually attributed to beliefs by common sense (1988, pp. 31-32)

But notice, to use this alternative account of reference to argue that beliefs and desires do exist is simply to substitute one argument from reference for another. This all suggests that in order to settle the ontological issues (e.g., do beliefs exist?), we need to determine whether the correct theory of reference is descriptivist or causal-historical.

The ambiguity theory of reference provides a new way of understanding some of the ontological disputes. Assume that much of the associated description surrounding the word ‘belief’ is pretty seriously mistaken. What should we conclude about the sentence “Beliefs exist”? Stephen Stich quite explicitly adopts a description theory of reference (1983, 17-23) and argues that there is no such thing as belief. Lycan, by contrast, explicitly adopts a causal-historical theory (1988, 31-32) and insists that there is such a thing as belief. The ambiguity theory allows us to say that both of them are right. The right way to interpret Stich’s eliminativist claim “Beliefs don’t exist” is in terms of a descriptivist theory, so his utterance is true; the right way to interpret Lycan’s preservationist claim “Beliefs exist” is in terms of his self-avowed causal-historical commitment, and so his utterance is true too.³⁵ Each is right given the appropriate disambiguation for their claims. More generally, for a given utterance of the sentence “Beliefs exist”, whether it’s true depends on the appropriate disambiguation of “beliefs”

³⁵ To be clear: causal-historical theories do not guarantee existence since the use of a term may have never been grounded in a bona fide referent to begin with. (See Devitt 1981; Devitt and Sterelny 1999 for discussion of grounding.) Our aim is simply to illustrate how the posited ambiguity can help diagnose certain philosophical debates.

in the context. For some utterances of such a sentence, the appropriate way to fix reference is descriptivist, and for other utterances, the appropriate reference fixing relation is causal historical. It follows from this that in some contexts, it will be true to utter the sentence “Beliefs exist” and in other contexts it will be false to utter the same sentence.³⁶ Despite the paradoxical air, this conclusion is driven by the fact that kind terms are ambiguous, and so there is no real contradiction, any more than there is in the fact that the sentence “I have money in the bank” is true under one disambiguation and not under the other.

The fact that Lycan tells us that he intends his term “belief” to be interpreted in a causal-historical fashion is a very good clue that the convention that fixes reference in that case is causal-historical. Unfortunately, most cases aren’t as straightforward as this because – most people don’t come out and tell you “this is the theory that fixes the reference of my term”. But there are features of the discourse that can convey which convention determines reference. Just saying “Witches don’t exist” indicates that the operative convention is descriptivist, because that allows for the charitable interpretation on which the claim is not obviously false. Whereas if someone says “Although scientists have been wrong about almost every aspect of the atom, atoms are real”, it seems that the operative convention is causal-historical, since that allows for the charitable interpretation on which the claim is not obviously false.

Recognizing intra-individual ambiguity also illuminates a related issue we have long found intriguing, provocative, ... even alarming! In his essay, “Deconstructing the Mind”, Stich points out that if we look at historical examples of mistaken theories, we find that in some cases -

³⁶ We are using “sentence” in the colloquial way that allows for the possibility of a given sentence can have different interpretations. This usage is sometimes labeled “surface sentence”.

phlogiston, caloric, witches - our intellectual ancestors concluded that the posits don't exist; however in other cases - stars, atoms, and planets - our ancestors concluded that the posits did exist despite the mistaken beliefs about them (Stich 1996, p. 63). While we might hope that there are some principles of "rational ontological inference" that explain why it is right to say that atoms exist but phlogiston doesn't, Stich is skeptical that we will find any such principles that can do the needed work (pp. 63ff). Rather, he suggests that political and social factors play a crucial role in generating ontological conclusions:

In some situations, it is easier to get a grant or a promotion or to enhance one's reputation in the scientific community by announcing the discovery of a new entity or denying the existence of one previously claimed to exist. In other situations, it is more politically expedient to conclude that entities of a certain sort don't have some of the properties previously attributed to them and that experimental results or other phenomena can best be explained by attributing some rather different properties to those entities. Which conclusion the scientific community ultimately accepts may well be determined, in some cases, by factors like these (Stich 1996, p. 68).

What's more, Stich argues, endorsing such conclusions is rational (1996, p. 70). What makes Stich's proposal so shocking is that it suggests that politics actually determines ontology, and that is a very radical view indeed.

We think that the ambiguity theory of reference provides the resources for a more moderate way of accommodating Stich's observation that politics seems to play a role in determining which ontological claims are true. It's not that politics determines what exists. That is, it's not that the ontology is indeterminate until socio-political factors settle the matter. Rather,

flexibility enters the picture via the ambiguity of kind terms. On the ambiguity theory, in cases of contested kinds like phlogiston, ether, and gene, the kind terms are ambiguous such that there is a liberal interpretation on which the kind term refers to something and also a restrictive interpretation on which the kind term fails to refer. Because such ambiguity is (or at least can be) intra-individual, in some cases, many of us might have access to both interpretations.³⁷ That is where political and social factors can enter the picture. What politics can do is to establish consensus around one of these interpretations.

If reference really does play a central role in ontological disputes, then the ambiguity theory suggests that we need to be attentive to the role of reference conventions even in ontological disputes that have been resolved. An explanation for why we endorse or reject an ontological claim will often involve an explanation for why a certain referential convention was adopted. For instance, in contemporary culture, we endorse the claim “Witches do not exist”. But part of what funds that endorsement is that our ancestors reached consensus around one reference convention as opposed to some other convention.

In addition, our thesis suggests some concrete ideas about how disputes concerning ontology should be carried out. Participants to such disputes should not only try to make explicit the assumptions about reference invoked, but they should be open to the possibility that more than one theory of reference can be correct. It is appropriately conciliatory to acknowledge the availability of other conventions when disputing about ontological claims. But it is overly

³⁷ See also Mallon 2006.

concessive to accede to ontological verdicts that imperialistically presuppose that there is only one correct convention about how kind terms refer.³⁸

References

- Adams, Robert 1979: ‘Divine command ethics modified again’. *Journal of Religious Ethics* 7, pp. 66 – 79.
- Andreasen, Robin 2000: ‘Race: Biological reality or social construct?’ *Philosophy of Science* 67, pp. S653 – S666.
- Appiah, K. A. 1995: ‘The Uncompleted Argument: Du Bois and the Illusion of Race’. In Bell and Blumenfeld 1995, pp. 59-78.
- Bartholomaeus, Anglicus 1482: *De proprietatibus rerum*. Lyon: Nikolaus Philippi und Markus Reinhard.
- Bell, Linda and David Blumenfeld (eds.) 1995: *Overcoming Racism and Sexism*. Lanham, MD: Rowman and Littlefield.

³⁸ We are grateful to Dave Chalmers, Mike Dacey, Michael Devitt, Steve Downes, James Genone, Steven Gross, Philipp Koralus, Victor Kumar, Genoveva Marti, Steve Stich, Hannah Tierney, and audiences at the University of Barcelona, the University of California at Davis, the University of Kansas, and Washington University in St. Louis for discussion and comments on earlier versions of this paper. We are also grateful to three referees and an editor at *Mind* for helpful comments. Special thanks to Mike Bishop and Edouard Machery for extensive feedback on this work.

- Bezuidenhout, Anne and Marga Reimer, (eds.) 2004: *Descriptions and Beyond: An Interdisciplinary Collection of Essays on Definite and Indefinite Descriptions and Other Related Phenomena*. Oxford University Press.
- Bianchi, Andrea (ed.) forthcoming: *On reference*, ed.. Oxford: Oxford University Press.
- Bishop, Michael, and Stephen Stich 1998: 'The flight to reference, or how not to make progress in the philosophy of science'. *Philosophy of Science* 65: 33 – 49.
- Boyd, Richard 1983: 'On the current status of the issue of scientific realism'. *Erkenntnis*, 19: 45 – 90.
- 1988: 'How to be a moral realist'. In Sayre-McCord 1988.
- 2002: 'Scientific Realism'. *The Stanford Encyclopedia of Philosophy* (Summer 2002 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2002/entries/scientific-realism/>>.
- Braun, David 1993: 'Empty Names'. *Nous* 27: 449 – 469.
- Burge, Tyler. 1973: 'Reference and proper names'. *The Journal of Philosophy* 70: 425 – 439.
- Büring, Daniel forthcoming: 'Pronouns'. In Heusinger et al forthcoming.
- Churchland, Paul 1981: 'Eliminative materialism and the propositional attitudes'. *The Journal of Philosophy* 78, no. 2: 67-90.
- Cumming, Samuel 2008: 'Variabilism'. *Philosophical Review* 117, no. 4: 525-554.
- Deutsch, Max 2009: 'Experimental Philosophy and the Theory of Reference'. *Mind & Language* 24: 445 – 66.
- Dever, Joshua 1991: *Variables*. Dissertation. University of California, Berkeley.
- Devitt, Michael 1981: *Designation*. New York: Columbia University Press.

- 2004. 'The Case for Referential Descriptions'. In Bezuidenhout and Reimer, pp. 280-305.
- 2011. 'Experimental Semantics'. *Philosophy and Phenomenological Research*. 82: 418 – 35.
- Devitt, Michael & Kim Sterelny 1999: *Language and Reality: An Introduction to the Philosophy of Language*, 2nd Edition. MIT Press.
- Donnellan, Keith S. 1966: 'Reference and Definite Descriptions'. *The Philosophical Review* 75, no. 3: 281-304.
- 1970: 'Proper Names and Identifying Descriptions'. *Synthese* 21, no. 3: 335-358.
- Dretske, Fred 1983: *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.
- Evans, Gareth 1973: 'The Causal Theory of Names'. *Proceedings of the Aristotelian Society*, Supplementary Volume 47: 187 – 208. Reprinted in *Collected Papers* (Oxford University Press: 1985), page references to reprinted version.
- Fodor, Jerry 1987: *Psychosemantics*. Cambridge, MA: MIT Press.
- 1990: *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.
- Frege, Gottlob 1893/1952: 'On Sense and Reference', in P. Geach and M. Black (eds.), *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell.
- Frisch, Deborah 1993: 'Reasons for Framing Effects'. *Organizational Behavior and Human Decision Processes* 54: 399 – 429
- Geach, Peter 1969: 'The Perils of Pauline'. *The Review of Metaphysics* 23: 287 – 300.
- Genone, James, & Lombrozo, Tania 2012: 'Concept Possession, Experimental Semantics, and Hybrid Theories of Reference'. *Philosophical Psychology* 25: 717 – 742.
- Gordon, Robert 1986: 'Folk Psychology as Simulation'. *Mind & Language*, 1, 158-71

- Guttenplan, Samuel (ed.) 1994: *A Companion to the Philosophy of Mind*. Blackwell.
- von Heusinger, Klaus, Claudia Maienborn and Paul Portner, (eds.) forthcoming: *Semantics: An International Handbook of Natural Language Meaning*. Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science. Berlin: Mouton De Gruyter.
- Ichikawa, Jonathan, Ishani Maitra, and Brian Weatherson 2012: 'In Defense of a Kripkean Dogma'. *Philosophy and Phenomenological Research* 85, no. 1: 56-68.
- Isidore of Seville 2006: *The Etymologies of Isidore of Seville*. Trans. by S. Barney, W. Lewis, J. Beach, & O. Berghof. Cambridge: Cambridge University Press.
- Jylkkä, Jussi, Henry Railo, and Jussi Haukioja 2009: 'Psychological essentialism and semantic externalism: Evidence for externalism in lay speakers' language use'. *Philosophical Psychology* 22: 37 – 60.
- Kitcher, Philip 1978: 'Theories, Theorists and Theoretical Change'. *The Philosophical Review*: 519-547.
- 1993: *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford University Press.
- Kripke, Saul 1980: *Naming and Necessity*. Cambridge: Harvard University Press.
- 1977: 'Speaker's Reference and Semantic Reference'. *Midwest studies in philosophy* 2, no. 1: 255-276.
- Kuhn, Thomas 1970: *The Structure of Scientific Revolutions*, 2nd edition. Chicago: University of Chicago Press.

- Lam, Barry 2010: 'Are Cantonese Speakers Really Descriptivists? Revisiting Cross-Cultural Semantics'. *Cognition* 115: 320 – 9.
- Lasnik, Peter 1993: 'Existence Presuppositions and Background Knowledge'. *Journal of Semantics* 10: 113 – 122.
- Leslie, Sarah-Jane forthcoming: 'Essence and Natural Kinds'. In J. Knobe et al., *Oxford Studies in Experimental Philosophy*. Oxford: Oxford University Press.
- Lewis, David 1972: 'Psychophysical and Theoretical Identifications'. *Australasian Journal of Philosophy* 50: 249 – 258.
- 1994: 'Reduction of Mind'. In Guttenplan 1994.
- 1999: 'Many, But Almost One'. In *Papers in Metaphysics and Epistemology*. Cambridge University Press.
- Ludwig, Kirk 2007: 'The Epistemology of Thought Experiments: First Person Versus Third Person Approaches'. *Midwest Studies in Philosophy* 31, no. 1: 128-159.
- Lycan, William 1988: *Judgement and justification*. Cambridge University Press.
- 2008: *Philosophy of Language: A Contemporary Introduction*. New York: Routledge.
- Machery, Edouard, Max Deutsch, and Justin Sytsma forthcoming a: 'Speaker's Reference and Cross-Cultural Semantics'. In Bianchi forthcoming.
- Machery, Edouard., Christopher Olivola, Hyundeuk Cheon, Irma Kurniawan, Carlos Mauro, Noel Struchiner, and Harry Susianto forthcoming b: 'Is Folk Essentialism a Fundamental Feature of Human Cognition? A Cross-Cultural Study'.
- Machery, Edouard, Ron Mallon, Shaun Nichols, and Stephen Stich 2004: 'Semantics Cross-Cultural Style'. *Cognition* 92: B1 – B12.

- Machery, Edouard., Christopher Olivola and Molly De Blanc 2009: 'Linguistic and Metalinguistic Intuitions in the Philosophy of Language'. *Analysis*, 69, pp. 689 – 694.
- Mackie, John 1977: *Ethics: Inventing right and wrong*. New York: Penguin.
- Mallon, Ron 2006: 'Race: Normative, not metaphysical or semantic'. *Ethics* 116 (3): 525 – 551.
- Mallon, Ron, Edouard Machery, Shaun Nichols, and Stephen Stich 2009: 'Against Arguments from Reference'. *Philosophy and Phenomenological Research* 79: 332 – 356.
- Martí, Genoveva 2009: 'Against Semantic Multi-Culturalism'. *Analysis* 69: 42 – 8.
- Pliny the Elder 2009: *Natural History*, Books 8-11. Trans. H. Rackham. Cambridge, MA: Loeb Classical Library.
- Putnam, Hilary 1975: 'Explanation and Reference'. In *Mind, Language and Reality*. Cambridge: Cambridge University Press, pp. 196 – 214
- Reinhart, Tanya 1981: 'Pragmatics and Linguistics: An Analysis of Sentence Topics'. *Philosophica* 27, pp. 53 – 94.
- Russell, Bertrand. 1919: 'Knowledge by Acquaintance and Knowledge by Description'. In *Mysticism and Logic*: London: George Allen and Unwin.
- Salmon, Nathan 1986: *Frege's Puzzle*. MIT Press.
- 1998: 'Nonexistence'. *Nous* 32, pp. 277 - 319
- Sayre-McCord, Geoffrey (ed.) 1988: *Essays on Moral Realism*. Ithaca: Cornell University Press.
- Searle, John 1958: 'Proper Names'. *Mind* 67, pp. 166-173.
- 1983: *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

- Soames, Scott 2002: *Beyond Rigidity: The Unfinished Semantic Agenda of 'Naming and Necessity'*. Oxford University Press.
- Stampe, Dennis 1977: 'Toward A Causal Theory Of Linguistic Representation'. *Midwest studies in philosophy*. 2, pp. 42 – 63.
- Stanovich, Keith. 1999. *Who Is Rational?* Mahwah, NJ: Erlbaum.
- Stich, Stephen 1983: *From Folk Psychology to Cognitive Science*. MIT Press.
- 1996: *Deconstructing the mind*. Oxford University Press, USA.
- Strawson, Peter 1950: 'On Referring'. *Mind* 59, pp. 320 – 344.
- 1964: 'Identifying Reference and Truth-Values'. *Theoria* 30, pp. 96 – 118.
- Sytsma, Justin, and Jonathan Livengood 2011: 'A New Perspective Concerning Experiments on Semantic Intuitions'. *Australasian Journal of Philosophy* 89, pp. 315-32.
- Tversky, Amos, and Daniel Kahneman 1981: 'The Framing of Decisions and the Psychology of Choice'. *Science* 211, no. 4481, pp. 453-458.
- Von Fintel, Kai 2004: 'Would You Believe It? The King of France is Back!' In Bezuidenhout and Reimer 2004, pp. 315-341.
- Zack, Naomi 1993: *Race and Mixed Race*. Philadelphia: Temple University Press.