Routledge
Taylor & Francis Group

Concept possession, experimental semantics, and hybrid theories of reference

James Genone and Tania Lombrozo

Contemporary debates about the nature of semantic reference have tended to focus on two competing approaches: theories which emphasize the importance of descriptive information associated with a referring term, and those which emphasize causal facts about the conditions under which the use of the term originated and was passed on. Recent empirical work by Machery and colleagues suggests that both causal and descriptive information can play a role in judgments about the reference of proper names, with findings of cross-cultural variation in judgments that imply differences between individuals with respect to whether they favor causal or descriptive information in making reference judgments. We extend this theoretical and empirical line of inquiry to views of the reference of natural and nominal kind concepts, which face similar challenges to those concerning the reference of proper names. In two experiments, we find evidence that both descriptive and causal factors contribute to judgments of concept reference, with no reliable differences between natural and nominal kinds. Moreover, we find evidence that the same individuals' judgments can rely on both descriptive and causal information, such that variation between individuals cannot be explained by appeal to a mixed population of "pure descriptive theorists" and "pure causal theorists." These findings suggest that the contrast between descriptive and causal theories of reference may be inappropriate; intuitions may instead support a hybrid theory of reference that includes both causal and descriptive factors. We propose that future research should focus on the relationship between these factors, and describe several possible frameworks for pursuing these issues. Our findings have implications for theories of semantic reference, as well as for theories of conceptual structure.

Keywords: Concepts; Experimental Philosophy; Semantics; Theories of Reference

James Genone is a post-doctoral fellow in the Introduction to the Humanities Program at Stanford University. Tania Lombrozo is an Assistant Professor of Psychology at the Univeristy of California, Berkeley. Correspondence to: James Genone, Stanford University, Stanford, USA. Email: jgenone@stanford.edu

1. Introduction

In discussions of semantic reference, two theoretical approaches have tended to dominate the literature. Descriptive theories of reference propose that reference is determined by which object is picked out by descriptive information associated with a referring term. Causal theories of reference hold that which object a referring term picks out depends on a particular use of the term being passed on via a causal chain between the members of a linguistic community. Philosophical arguments on both sides of this debate have often relied on intuitions about hypothetical cases where the descriptive information associated with a referring term is incomplete or false of the intended referent. Many philosophers have taken the intuitive force of these examples to support the causal theory of reference. Recently, however, some philosophers (Machery, Mallon, Stich, & Nichols, 2004; Mallon, Machery, Nichols & Stich, 2009) have argued that the methodology of basing arguments on intuitions about particular cases is unreliable, because although there may be considerable consensus among philosophers about such cases, experiments show that a similar consensus is not found among non-philosophers, particularly when one compares different cultural groups.

In this paper, we aim to shed new light on these issues. To begin with, we discuss how the debate between causal and descriptive theories of reference arises in the domain of concepts, rather than just that of proper names (section 2). Second, we explain why a hybrid view of reference, which incorporates both causal and descriptive factors as relevant to reference determination, should be part of the discussion, in addition to pure causal and descriptive theories (section 3). Third, after discussing recent empirical research (sections 4), we present our own empirical results suggesting that individuals utilize both causal and descriptive information in making judgments about the reference of concepts from a variety of domains (sections 5 and 6). We then argue that the extant empirical literature does not by itself undermine the practice of supporting theories of reference with arguments that appeal to intuitions about hypothetical cases. What such intuitions may suggest is that the correct theory of reference needs to incorporate both causal and descriptive factors (section 7). We conclude by discussing ways in which future research in this area might aim to discover the mechanisms governing our use of causal and descriptive information in making judgments about reference (section 8).

2. Concepts and Reference

Many theories of concepts propose that grasp of a concept depends on possessing correct descriptive information about the concept's referent—that grasping the concept GOLD, for example, is a matter of believing that gold is a shiny, malleable metal often used for making jewelry. People often lack extensive knowledge of the properties and kinds picked out by even the most familiar concepts, however, and frequently possess false beliefs about them, making it difficult for theories that emphasize descriptive information to explain even the most ordinary cases of

concept possession. This challenge is sometimes called *the problem of ignorance and error* (Laurence & Margolis, 1999). One solution to the problem would be to reject the idea that there is any particular descriptive content that one must associate with a concept; rather, concepts are assigned contents depending upon the inferential dispositions of particular individuals. On this view, one can grasp a given concept despite the descriptive content associated with it being incomplete or even incorrect. This view faces a different problem, however: if concepts are idiosyncratic and differ significantly in their content across individuals, in virtue of what can we say that two people share a particular concept?

A problem similar to that of ignorance and error arises in semantics for certain theories of the reference of proper names. According to *descriptivist* accounts, the referent of a proper name is determined by the descriptive content associated with that name (Frege, 1892/1948; Searle, 1958). If we imagine that the description commonly associated with a proper name is false, for example if Gödel stole his famous incompleteness theorem from someone named 'Schmidt', as in Saul Kripke's (1972/1980) famous example, then according to the descriptive theory the name 'Gödel' will counter-intuitively pick out Schmidt, since he is the one who fulfills the description 'prover of the incompleteness theorem'. In *Naming and necessity*, Kripke argued that rejecting descriptivism and instead adopting a view of reference that emphasizes causal connections provides a more intuitive answer to the question of whom the name 'Gödel' refers to. Roughly, Kripke's causal view claims that the reference of a name is secured by a causal chain linking users of the name with an initial baptizing event in which the name is associated with the particular person, place, or thing it names. Gödel gets his name from his parents, and everyone else is able to refer to him using that name by being part of a causal chain that links each person to the initial use of the name introduced by Gödel's parents (Kripke, 1972/1980).[1] So long as one is part of the right causal chain, one can use the name 'Gödel' to refer to Gödel regardless of how little information about him one possesses, or how much of it is false.

Kripke's view, which has come to be known as the *causal theory of reference for proper names*, is generally thought to be an advance over descriptivism, and philosophers have argued that its application is not restricted to proper names, but that the basic idea can be extended to natural kind terms (Putnam, 1975) and perhaps to a wide variety of concepts (Burge, 1979). Burge's work is particularly suggestive with respect to the problem of ignorance and error for concepts mentioned above—he argues that we are inclined to attribute possession of a concept to individuals despite significant ignorance or false beliefs as long as they intend to use the concept with the same reference as those in their linguistic community who have adequate mastery of the concept.[2] For example, Burge describes a case in which a patient with arthritis in his joints falsely believes that it has spread to his thigh. When the patient is corrected by his doctor, however, he revises his belief in deference to his doctor's expertise. Burge argues that given the patient's disposition to defer to those with greater knowledge, we should interpret the patient's original belief as involving the concept of arthritis despite his false belief that the

disease can spread from joints to muscles (Burge, 1979). Burge's argument suggests that just as we have reason to accept Kripke's causal theory for the reference of proper names, we should also accept a *causal theory of reference for concepts* in order to avoid the problem of ignorance and error. Just as failure to possess correct or sufficient descriptive knowledge of the referent of a name does not prevent an individual from using that name if she is part of a causal chain that preserves the use of the name, an individual can also be said to possess a concept despite considerable ignorance or false information about it if she is appropriately causally connected to expert users of the concept.[3]

3. Hybrid Reference

Despite the popularity of the causal theory of reference, philosophers have since demonstrated that the view stands in need of refinement. Evans (1973), for example, generated counter-examples to Kripke's causal view by thinking about cases of reference change—cases where the object referred to with the use of a name changes unbeknownst (at least initially) to users of the name. Evans described a case involving Marco Polo: when Marco Polo learned the name 'Madagascar' from local sailors, he intended to use it with the same reference with which they used it. He mistakenly believed, however, that they were referring to the island that now bears the name, rather than a portion of the mainland that was the (then) true referent. Given that Marco Polo intended to use the name 'Madagascar' in the same way as the local sailors from whom he learned it, Kripke's causal view predicts that his usage (and all subsequent uses deriving from him) would refer to the mainland rather than the island. In defense of Kripke's view we might say that Marco Polo's intention to refer to what the locals were referring to was unsuccessful, and this is what explains why *his* use of 'Madagascar' refers to the island rather than the mainland. This doesn't explain the change in the reference of the name, however. For that, one would on Kripke's theory expect a new baptizing event, but Marco Polo's initial mistaken usage doesn't seem to count as such an event given that he didn't intend to use the name with a new reference.

Evans suggested that in order to deal with this and other problematic cases, descriptive information must play a role in reference determination. He proposed that each use of a name is associated with a file of information that an individual stores about the referent of the name. While some of the information in the file can be false (by not being generated by the appropriate source), reference will be tied to whatever is the dominant causal source of the information in the file. In the case of Madagascar, since over time the majority of subsequent users' files associated with the name contained information about the island (since Marco Polo's error wasn't corrected), the island became the new referent of the name. Evans believed that a theory which incorporated *both* causal and descriptive factors in determining reference would best explain the reference of proper names, as opposed to the pure causal or descriptive theories that had been previously favored.[4]

Evans's proposal raises the possibility of a *hybrid theory of reference*: one that incorporates both descriptive and causal factors as relevant to determining reference. There are many ways in which a hybrid theory of reference might be formulated, and in particular different ways in which causal relations and descriptive information associated with a referring term might work together or independently in order to pick out a referent. According to some views, the causal source of some set of descriptions associated with a proper name will pick out the referent of the name. On Evans's view, for example, the referent is the source of the majority of the associated descriptions, but the relevant set could be picked out in other ways.[5] Alternatively, facts about causal relations and descriptions associated with the use of a proper name might function as independent conditions on reference, such that a successful user of a name might need to both stand in the appropriate causal relations and associate some relevant set of descriptions with the name.

For present purposes, we won't focus on the different forms a hybrid theory might take, since our aim in presenting this proposal is simply to motivate consideration of a hybrid theory as a possible alternative to pure causal or descriptive theories, and in particular as a useful framework for empirical work on theories of reference. In sections 5 and 6, we present empirical evidence that individuals' intuitions about reference depend on both causal and descriptive factors, and go on to argue that this undermines support for either a pure causal theory or a pure descriptive theory of reference. In section 7, we will consider what constraints our results might place on different versions of a hybrid theory of reference. Before moving on to our experiments, however, we will introduce previous work in experimental semantics that has aimed to identify the extent to which folk intuitions conform to either causal or descriptive theories, and how such judgments vary across populations.

4. Experimental Semantics

In a widely discussed paper, Edouard Machery and colleagues (Machery et al., 2004) note that in philosophical literature semantic theories are often motivated by appealing to intuitions about hypothetical cases, such as the Gödel case described above, revealing a methodological commitment to the idea that ''theories of reference for names have to be consistent with our *intuitions* regarding who or what the name refers to'' (Machery et al., 2004, p. B2). Machery et al. then present empirical findings which they take to make a prima facie case for cross-cultural variation among non-philosophers with respect to intuitions about whom a proper name refers to when the user of the name has false beliefs about the bearer of the name. These findings challenge the idea that intuitions about hypothetical cases, even if widely shared by philosophers, can be used in support of a particular theory about the reference of proper names.

Machery et al.'s experimental design involved presenting American and Chinese participants with vignettes involving false descriptive information associated with a proper name followed by a question about the reference of the name. One of the

vignettes and reference questions, which was based directly on Kripke's Gödel case, ran as follows:

> Suppose that John has learned in college that Gödel is the man who proved an important mathematical theorem, called the incompleteness of arithmetic. John is quite good at mathematics and he can give an accurate statement of the incompleteness theorem, which he attributes to Gödel as the discoverer. But this is the only thing that he has heard about Gödel. Now suppose that Gödel was not the author of this theorem. A man called ''Schmidt,'' whose body was found in Vienna under mysterious circumstances many years ago, actually did the work in question. His friend Gödel somehow got hold of the manuscript and claimed credit for the work, which was thereafter attributed to Gödel. Thus, he has been known as the man who proved the incompleteness of arithmetic. Most people who have heard the name 'Gödel' are like John; the claim that Gödel discovered the incompleteness theorem is the only thing they have ever heard about Gödel.

> **When John uses the name 'Gödel', is he talking about:**

> (a) the person who really discovered the incompleteness of arithmetic? or
> (b) the person who got hold of the manuscript and claimed credit for the work?
> (Machery et al. 2004, p. B6)

Answers of (B) to the reference question, which are consistent with descriptivist theories of reference, were given a score of 0 and answers of (A)—consistent with a causal account—were given a 1. The scores for each of two similar vignettes were summed, resulting in a cumulative score ranging from 0 to 2. The means (with standard deviations in parentheses) were:

> American participants: 1.13 (0.88)
> Chinese participants: 0.63 (0.84)

In Machery et al.'s study, Chinese participants interpreted the reference of 'Gödel' in the story in a manner more consistent with a descriptivist account of reference, while the American participants' answers were more consistent with a causal interpretation of reference. Nevertheless, it is important to note that although the two populations differed in tending to favor judgments consistent with different theories of reference, there was also significant variation within each population, with some Americans responding in a way consistent with a descriptivist theory and some Chinese participants in a way consistent with a causal theory.[6] So although Machery et al. highlight the cross-cultural divergence in intuitions about reference by comparing their data to evidence of cross-cultural variation in cognitive styles discussed in the social psychology literature, their findings also suggest significant variation within a given culture.[7]

Machery et al.'s findings are consistent with the proposal that intuitions about reference can be sensitive to either causal or descriptive factors, since according to their data individuals differ with respect to which theory their intuitions favor. This kind of variation across individuals is potentially unsurprising given that philosophers show corresponding variation, with some endorsing descriptive theories and others causal theories.[8] A hybrid theory of reference, however, predicts use of both

factors in making reference judgments *by the same individual*, rather than differential use across individuals, which is what Machery et al.'s findings establish. Accordingly, one of our aims in the experiments described below is to examine internal consistency in individuals' intuitions about reference to determine whether individuals' judgments can be appropriately characterized as strictly descriptivist or strictly causal. To the extent that individuals' judgments are a function of both factors, categorizing individuals' intuitions as either purely descriptive or purely causal may misrepresent the nature of their actual judgments, and obscure the relationship between folk judgments and philosophical debates about reference. Importantly, for Machery and colleagues to undermine the practice of relying on intuitions to support theories of reference, it would need to be the case that the variation in intuitions they report could not be accounted for by a hybrid theory of reference. After presenting evidence that individuals do not have purely causal or descriptive intuitions, we will suggest how a hybrid theory of reference might be able to account for Machery et al.'s results.

A second aim of the experiments we present is to investigate intuitions about reference for concepts rather than for proper names, including concepts of natural kinds (diseases, minerals) and nominal kinds (artifacts, legal documents). Examining how information about false beliefs and causal factors affects participants' willingness to attribute possession of such concepts to a third party has implications for whether a hybrid theory of reference can provide a solution to the problem of ignorance and error. And by considering different domains, the experiment also bears on debates about whether causal approaches to concepts apply only to natural kinds (such as diseases and minerals) or to other kinds as well (such as artifacts and legal documents, which are typically regarded as nominal kinds; Devitt & Sterelny, 1999, chapter 5).

5. Experiment 1

In experiment 1, participants were asked whether or not the concepts two characters in a vignette associate with a common word share the same reference. There were four types of vignettes, constructed to cross the predictions of a descriptive theory with those of a causal theory. Thus the two characters could either have matching or mismatching descriptive information associated with the concept, and either be linked to a common causal chain or to different causal chains. Each participant provided judgments for all four types of vignettes, allowing us to examine consistency across judgments, and in particular whether participants tended to respond on the basis of a single factor, as might be expected if individuals are best characterized as either pure descriptive or causal theorists. We additionally varied the amount of mismatching descriptive information to examine whether this would have a graded effect on judgments concerning shared reference, and whether the concept in question was a natural kind or a nominal kind, in order to examine whether descriptive and causal factors contribute differentially to these two domains of concepts.

Participants. Participants were 192 undergraduates who completed a questionnaire for course credit along with unrelated experiments.

Materials and procedures. Participants learned about two fictional islands with agents whose beliefs about a concept either matched or mismatched those of experts. By varying whether beliefs matched, we could vary whether two people's concepts had the same referent according to descriptive theories. By varying whether two people were on the same island, we could vary whether their concepts had the same referent according to causal theories. This resulted in four possible judgments about reference (description same or different x causal origin same or different) for each participant.

The questionnaire had four parts. In part 1 (description: different, causal origin: same), participants learned about Alex from island Alpha. Below is a sample scenario:

> There is a small island in the Indian Ocean called Alpha. Natives of Alpha, called "Alphians," sometimes catch diseases not found anywhere else in the human population. When this happens, they consult Alphian doctors. One of the diseases on Alpha is tyleritis. Tyleritis is a disease that affects muscles and causes muscle pain. It is caused by exposure to a rare mineral, can be diagnosed with a blood test, and can be cured by an injection.
>
> **Facts about the Alphian disease called "tyleritis":**
>
> - Tyleritis affects the muscles and causes muscle pain.
> - Tyleritis is caused by exposure to a rare mineral.
> - Tyleritis can be diagnosed with a blood test.
> - Tyleritis can be cured by an injection.
>
> Alex is a native Alphian. He has recently felt pain in his joints. He knows that tyleritis is a disease, and that it can cause pain. Alex has a number of other beliefs about tyleritis. He thinks tyleritis can cause pain in the joints, and hence thinks he might have tyleritis. He also believes tyleritis is caused by exposure to a rare mineral, that it is diagnosed with a blood test, and that it is cured by an injection.
>
> **Alex's beliefs about the Alphian disease called "tyleritis":**
>
> - Tyleritis affects the joints and causes joint pain.
> - Tyleritis is caused by exposure to a rare mineral.
> - Tyleritis can be diagnosed with a blood test.
> - Tyleritis can be cured by an injection.
>
> Alex decides to consult his doctor to find out if he has the Alphian disease called "tyleritis." His doctor is an expert on the Alphian disease called "tyleritis." His doctor knows that tyleritis is a disease that affects muscles and causes muscle pain. He also knows that it is caused by exposure to a rare mineral, can be diagnosed with a blood test, and can be cured by an injection.
>
> **Alex's doctor's beliefs about the Alphian disease called "tyleritis":**
>
> - Tyleritis affects the muscles and causes muscle pain.
> - Tyleritis is caused by exposure to a rare mineral.
> - Tyleritis can be diagnosed with a blood test.
> - Tyleritis can be cured by an injection.

In the examination room, Alex and his doctor each have an immediate thought:

Alex thinks: ''I might have tyleritis.''
His doctor thinks: ''I wonder if this patient has tyleritis.''

When Alex and his doctor each have a thought in the examination room, are they having a thought about the same disease?

(A) Yes, they are both thinking about the same disease.
(B) No, they are not thinking about the same disease.

While all scenarios followed the same structure, the actual concept in question varied across conditions (disease, mineral, artifact, legal term), as did the number of Alex's false beliefs (see appendix A).

In part II (description: same, causal origin: same), participants learned about Bob from island Brom, where there is a different disease also called ''tyleritis'' found nowhere else. Bob and his doctor have identical and true beliefs. Participants were asked about shared reference for Bob and his doctor as in part I.

Critically, Bob's beliefs were always identical to Alex's, so in part III (description: same, causal origin: different), participants were asked whether Alex and Bob had the same referent when thinking about the disease each called ''tyleritis.'' Below is sample text:

Alex from Alpha and Bob from Brom had the following thoughts over breakfast before leaving for their respective doctors:

Alex: ''I hope I don't have tyleritis.''
Bob: ''I think I have tyleritis.''

When Alex and Bob each have a thought over breakfast, are they having a thought about the same disease?

(A) Yes, they are both thinking about the same disease.
(B) No, they are not thinking about the same disease.

Finally, in part IV (description: different, causal origin: different), participants learned that Alex and Bob both used a common word ('simus') involving different causal origins and different beliefs, and were asked about shared reference.

Parts I, II, and III each included additional true/false questions to ensure that participants understood the scenarios and appreciated which beliefs did and did not match in each case. For example, in part I participants were asked, ''would Alex and his doctor *both* agree with the following claims?'', and then presented with each possible belief (e.g., ''tyleritis affects the muscles and causes muscle pain''). Parts II and III had equivalent questions for the corresponding characters: Bob and his doctor in part II, and Alex and Bob in part III. There were eight possible beliefs in each part, for a total of 24 (8 questions x 3 parts) true/false questions per participant.

The order of parts I and II was counterbalanced, as was the order of the reference and true/false task in each part. Participants were randomly assigned to a condition involving one, two, three, or four false beliefs for Alex, and involving a concept from one of four domains (see appendix A for a complete set of stimulus materials).

Results and discussion. Before considering the primary questions of interest concerning judgments about reference, it is worth noting that performance on the true/false comprehension task was quite high, with a mean score of 22.26 (SD = 2.08) out of 24. Moreover, performance did not vary systematically as a function of the number of Alex's false beliefs, suggesting that differences in reference judgments across conditions are not attributable to differences in comprehension.[9]

Table 1 reports the percent of participants responding "yes" to the question of shared reference in each part of the questionnaire. Importantly, participants overwhelmingly ascribed shared reference to the agents in part II and not in part IV; these are the cases for which descriptive and causal theories agree. Responses for judgments pitting predictions of descriptive and causal theories (parts I and III) were intermediate, and did not differ significantly from 50% (I: $\chi^2(1) = 1.26$, $p = 0.26$; III: $\chi^2(1) = 0.38$, $p = 0.54$) or from each other ($\chi^2(1) = 3.01$, $p = 0.08$), although there was a trend towards greater agreement with shared reference in part III.

We next examined judgments as a function of the number of Alex's false beliefs. On part I, responses did vary as a function of false beliefs, $\chi^2(3) = 29.98$, $p < 0.01$, with fewer participants endorsing shared reference as the number of false beliefs increased (see figure 1). This suggests that the response rate near 50% for part I does not reflect an absence of intuitions on the part of participants, as responses did vary systematically in response to the amount of overlapping descriptive information. On part III, responses did not vary significantly as a function of the number of false beliefs, $\chi^2(3) = 3.43$, $p = 0.33$.

These findings suggest that participants are sensitive to both descriptive and causal information. This could arise because most participants are sensitive to both factors,

Table 1 Summary of Four Parts of Questionnaire in Experiments 1 & 2.[a]

	Example	Exp 1	Exp 2
Part I:			
Description: different	Alpha agent: "tyleritis curable"		
Causal origin: same	Alpha expert: "tyleritis incurable"	44%	4.44 (0.17)
Part II:			
Description: same	Brom expert: "tyleritis curable"		
Causal origin: same	Brom agent: "tyleritis curable"	98%	6.37 (0.11)
Part III:			
Description: same	Alpha agent: "tyleritis curable"		
Causal origin: different	Brom agent: "tyleritis curable"	53%	3.84 (0.20)
Part IV:			
Description: different	Alpha agent: "simus has P"		
Causal origin: different	Brom agent: "simus has Q"	2%	1.95 (0.18)

Notes: [a]Also indicated are the percent of participants responding "yes" to each question of shared reference (experiment 1) and the ratings of agreement with shared reference on a 7-point scale (experiment 2) with standard errors of the means in parentheses.
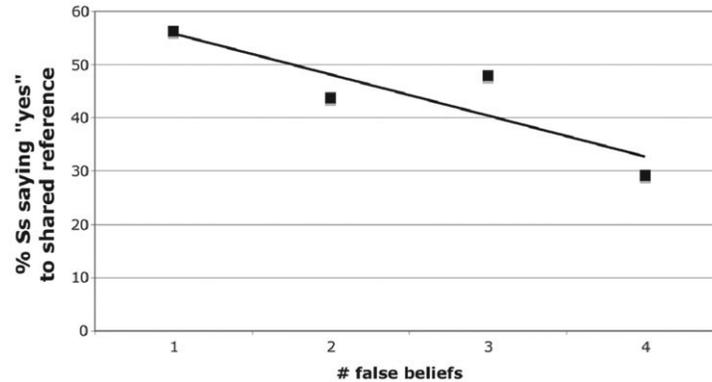
Figure 1. Percent of Participants Attributing Shared Reference for Part I of Experiment 1 as a Function of Alex's Number of False Beliefs.

or because different individuals have different intuitions or adopt different strategies: some descriptive and some causal. These possibilities can be distinguished by examining the relationship between participants' judgments on parts I and III. If individuals vary in their sensitivity to or preferential reliance upon these factors, with some closer to being "pure descriptive theorists" and others "pure causal theorists," a response of "yes" on part I (causal) should be correlated with a "no" on part III, and vice versa. In fact, these two judgments were not significantly correlated ($r = 0.081$, $p = 0.26$).[10] This suggests that most participants utilize both descriptive and causal information in making reference judgments rather than preferentially applying one factor or the other across judgments.

Finally, and perhaps surprisingly, domain did not significantly influence reference judgments in Part 1, $\chi^2(3) = 6.65$, $p = 0.08$, or in part III, $\chi^2(3) = 4.94$, $p = 0.18$ (see table 2). While it is difficult to draw strong conclusion from the failure to find a difference, we can at least conclude that our results do not provide positive support for the proposal that causal or descriptive factors contribute differentially to judgments about reference in different domains.

Table 2 Responses on Reference Questions for Experiments 1 and 2 as a Function of Domain.[a]

	Experiment 1		Experiment 2	
	Part I	Part III	Part I	Part III
Disease	46% (7.3)	56% (7.2)	4.81 (0.28)	4.41 (0.40)
Mineral	54% (7.3)	56% (7.2)	4.03 (0.34)	3.69 (0.39)
Artifact	29% (6.6)	40% (7.1)	4.41 (0.39)	3.56 (0.42)
Legal Doc	48% (7.3)	60% (7.1)	4.50 (0.34)	3.72 (0.35)

Note: [a]Means are followed by standard errors of the means in parentheses.

In sum, experiment 1 had two central aims: to examine the relationship between descriptive and causal factors in participants' judgments about shared reference, and to do so in the context of judgments about concepts as opposed to proper names. With respect to the first aim, the findings support the hypothesis that reference judgments are a function of both descriptive and causal factors. As a group, participants' judgments across the four parts of the task revealed very consistent responses when descriptive and causal factors pointed towards the same judgment (parts II and IV), and mixed responses when the factors were in conflict, suggesting that participants adopted distinct strategies for reconciling the conflict. And at the level of individuals, strategies for reconciliation did not take the form of consistent reliance on a single factor, as this would have generated a correlation between responses on parts I and III, which was not observed.

With respect to the second aim, we succeeded in extending experimental semantics to concepts, having identified two factors—namely, overlapping descriptive information and shared causal origin—as significant influences on folk judgments of shared reference for concepts from four domains. We did not find evidence of differential reliance on these factors across domains, though as noted, this null result does not support strong conclusions.

Experiment 2 aimed to replicate and clarify the interpretation of these initial, promising results.

6. Experiment 2

Experiment 2 sought to replicate and extend experiment 1. First, the experiment addressed a deficiency in experiment 1: much of the information relevant for applying a causal theory was unspecified (e.g., how Alex learned about tyleritis), making it possible that participants drew inferences beyond the stimulus materials (e.g., that Alex learned about tyleritis from listening to Bromian radio). In experiment 2, we manipulated whether participants received additional information specifying a learning event. Second, we were interested in whether having participants justify their reference judgments would induce greater within-subject consistency across parts I and III. Finally, we hoped to replicate the effect of number of false beliefs (one versus four), and to confirm the absence of sizeable domain effects.

Participants. Participants were 128 undergraduates who completed the questionnaire for course credit along with unrelated experiments.

Materials and procedures. The questionnaire paralleled experiment 1, with the following key modifications. (a) Participants in the *history specified* condition saw an additional sentence indicating how each agent learned about the relevant concept (e.g., "Alex first heard of tyleritis when his uncle contracted it and he overheard other family members discussing it"). (b) Participants in the *justification* condition received a prompt to explain after each reference judgment (e.g., "in a sentence or two, please explain why you think Alex and his doctor are or are not having thoughts about the same disease"). (c) Instead of a "yes" or "no" question about shared

reference, participants were asked to make a reference judgment on a 7-point scale by indicating their agreement with a claim about shared reference (e.g. "when Alex and his doctor each have a thought in the examination room, they are having thoughts about the same disease"). (d) The number of Alex's beliefs that mismatched those of the expert in part I were either 1 or 4, without the intermediate conditions included in experiment 1.

Participants were randomly assigned to one of 32 conditions, the result of crossing domain (4: disease, mineral, artifact, legal document), number of false beliefs (2: 1, 4), justification (2: present, absent), and history (2: specified, unspecified). As there were no order effects in experiment 1, parts I–IV were always presented in the same order. We also omitted the comprehension questions from experiment 1 given the near-ceiling performance across conditions.

Results and discussion. Replicating experiment I, participants overwhelmingly endorsed shared reference in part II and rejected it in part IV, with intermediate ratings for parts I and III (see table 1, figure 2, and figure 3). Ratings for part I were significantly higher than the mid-point of the scale ($M = 4.44$, SD $= 1.92$, $t(127) = 2.57$, $p < 0.05$) and than ratings on part III ($M = 3.84$, SD $= 2.22$, $t(127) = 2.23$, $p < 0.05$), which did not differ from the midpoint, $t(127) = -0.80$, $p = 0.43$.
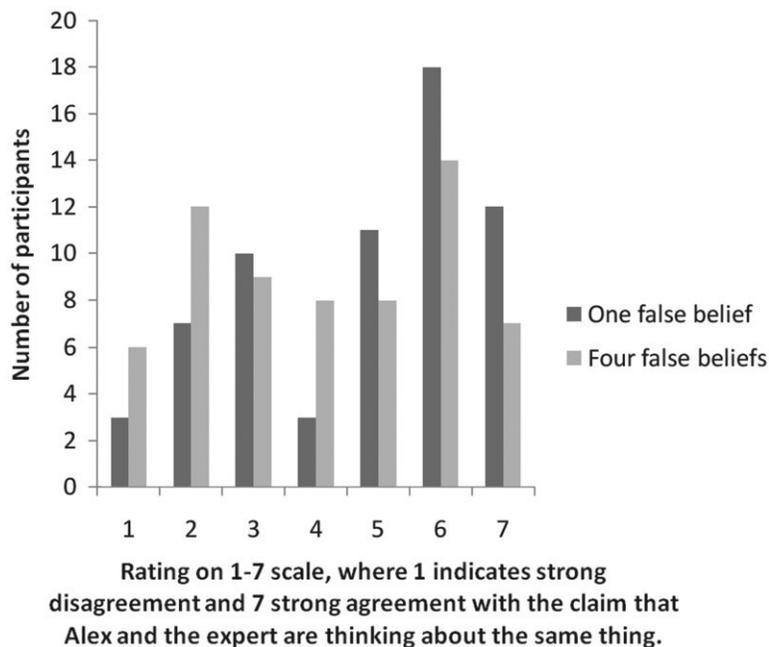

Figure 2. Distribution of Responses to The Shared Reference Question in Part I of Experiment 2 as a Function of Alex's Number of False Beliefs.
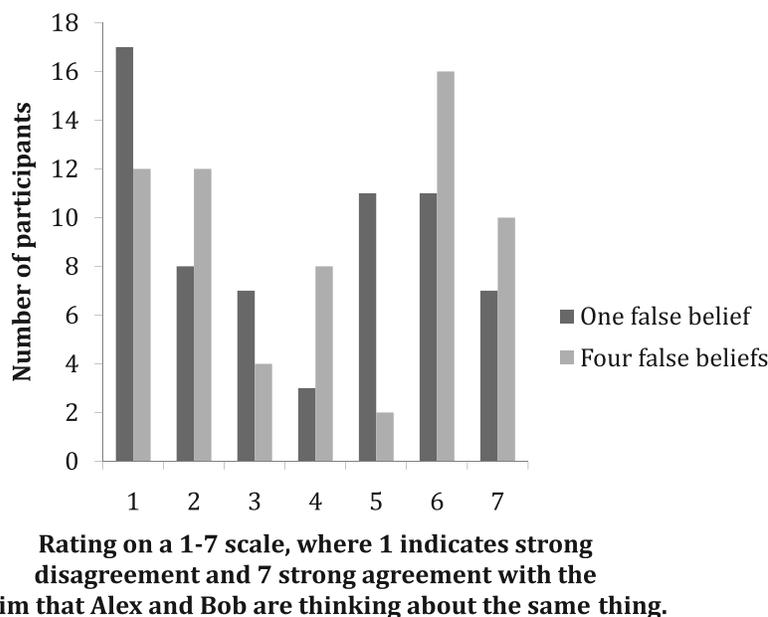
Figure 3. Distribution of Responses to the Shared Reference Question in Part III of Experiment 2 as a Function of Alex's Number of False Beliefs.

To further analyze responses on part I, we performed an ANOVA with domain (disease, mineral, artifact, legal document), number of false beliefs (1, 4) and history (specified, unspecified) as between-subjects variables, and agreement with shared reference as the dependent measure. (The justification variable was excluded, as participants provided their first justification only after part I judgments.) This analysis revealed a significant effect of false beliefs, $F(1,112) = 4.03$, $p < 0.05$. When Alex had one false belief, mean agreement with shared reference was 4.78 (SD = 1.86), but this number dropped to 4.09 (SD = 1.94) when he had four. There were no significant effects of domain, $F(3,112) = 0.883$, $p = 0.45$ (see table 2) nor of history, $F(1,112) = 0.133$, $p = 0.72$, nor were there significant interactions.

These findings replicate those for part I from experiment 1, with a reliable effect of number of false beliefs and no significant effect of domain. They also go beyond experiment 1 in testing the role of a specified history. That this did not reliably influence judgments suggests that the findings in experiment 1 were not an artifact of having failed to specify a learning history in the vignettes.

While mean ratings for shared reference were again near the middle of the scale (4.44 on a 1–7 scale), this was not because a majority of participants had weak or middling intuitions. Figure 2 reports the distribution of ratings on the 1–7 scale, and reveals what appears to be a bimodal distribution[11]: the mean of 4.44 resulted not from a majority of participants clustered around 4 and 5, but rather from a

combination of ratings near the bottom and top of the scale.[12] This suggests that participants had strong intuitions in response to the questions about shared reference.

To analyze reference judgments on part III, we repeated an ANOVA with the variables included for judgments from part I, but incorporating the variable of justification (2: present, absent). This analysis revealed a suggestive effect of history $F(1,96) = 2.86$, $p = 0.09$, but no other main effects (domain: $F(3,96) = 0.97$, $p = 0.41$, false beliefs: $F(1,96) = 0.65$, $p = 0.42$, justification: $F(1,96) = 0.53$, $p = 0.47$) nor interactions. When learning history was specified, participants provided lower ratings of shared reference in part III than when learning history was unspecified, 3.52 (SD = 2.10) versus 4.17 (SD = 2.29). If this effect is reliable, it suggests that while participants were sensitive to the history manipulation, it made shared descriptive information insufficient for shared reference (part III) rather than overriding mismatching descriptive information in the face of a common causal history (part I). The findings from judgments on part III otherwise replicate those from experiment I, with no significant effects of number of false beliefs or of domain.

Figure 3 reports the distribution of responses to the reference question for part III as a function of number of false beliefs. As with part I ratings, responses appear to be bimodal,[13] with a majority of participants providing high or low ratings, and few clustered at the scale mid-point. This suggests that participants had strong intuitions about the reference question, and were not guessing or responding at random.

In sum, the findings from the reference judgments for parts I and III are consistent with experiment 1, replicating effects of false beliefs with no sizeable domain effects. Also as in experiment 1, individual participants' judgments on parts I and III were not significantly related ($r = -.056$, $p = 0.53$)[14], suggesting that most individuals generated intuitions on the basis of both descriptive and causal information. However, the findings go beyond experiment 1 in a number of ways. First, we tested the hypothesis that specifying agents' learning histories would increase responses consistent with causal theories. This hypothesis was not supported, although specifying learning history did potentially lower the importance of shared descriptive information in part III. Second, we examined whether prompting participants to justify their reference judgments on part I would induce greater consistency. Justifying reference judgments pushed participants closer to the predicted negative correlation between parts I and III (no justification: $r < 0.01$, $p = 0.99$, justification: $r = -0.12$, $p = 0.22$), but the effect was small and not statistically significant. A negative result is hard to interpret, but it reinforces our suggestion that most individuals utilize both descriptive and causal information in making judgments about reference, and that mixed patterns do not reflect either the absence of intuitions among participants or a heterogeneous population with some "intuitive descriptive theorists" and some "intuitive causal theorists," as suggested by Machery et al.'s studies discussed above. Finally, experiment 2 employed a continuous dependent measure—a 1–7 point rating scale—rather than a dichotomous yes/no question. This allows us to consider more directly whether ratings near 50% or the scale mid-point reflect an absence of strong intuitions across participants, or instead

strong intuitions that differ across participants. The distributions of responses for the reference questions in both part I and part III (figures 2–3) support the latter interpretation: a majority of participants responded towards one end of the scale, not at the mid-point.

7. Discussion

The results obtained from these experiments suggest, in line with a hybrid theory of reference, that successful accounts of reference and concept possession will need to include both causal and descriptive factors. The fact that participants' responses are sensitive to manipulations of both causal and descriptive information, but not consistent across different conditions in privileging one sort of information over the other, suggests that both factors may play a role in individual reference judgments. Moreover, we have strong evidence that variation in responses did not result from weak or absent intuitions. First, the number of false beliefs in part I had a reliable impact on judgments, suggesting that participants were sensitive to the properties of the vignettes, and were not responding at random. Second, the distribution of responses to parts I and III in experiment 2 were bimodal, with few participants clustered at the scale midpoints.

Our findings suggest that the common practice of contrasting pure causal and pure descriptive theories of reference, which Machery et al. adopt, may be mistaken. As mentioned above, while their findings predict sensitivity to either causal or descriptive information, the contrast between causal and descriptive theories of reference along with the cross-cultural variation they report points to variation *between* individuals rather than *within* individuals.[15] The main aim of Machery and his collaborators is to criticize the reliance of philosophers on intuitions about thought experiments, and their findings are meant to show that such intuitions are not reliable. Showing that there is variation between cultural groups may only show, however, that intuitions don't favor *pure* causal or descriptive theories. A more sophisticated hybrid theory of reference, involving both descriptive and causal elements, might be able to explain the variation in folk intuitions, which would undermine the claim that such variation creates problems for the methodology of relying on intuitions, and instead suggest that intuitions support a hybrid theory of reference.[16]

One way in which a hybrid theory could explain the variation across cultures and across individuals within a culture would be if individuals differed, consciously or unconsciously, in their preferred strategy for combining causal and descriptive information in making reference judgments. Which strategy they applied could depend on various contextual factors, many of which are either unspecified or ambiguous in the vignettes participants are asked to evaluate. For example, participants may vary in the assumptions they make about the agent in the vignette's background beliefs or learning history. Depending on the additional assumptions participants make, they might privilege causal or descriptive information, thus giving answers more in line with one or the other of those theories. If this is correct, it may

suggest that the cultural differences uncovered by Machery et al. point to cultural differences in assumptions about missing details in the vignettes, rather than differences that reflect divergent intuitions about whether causal or descriptive factors are what determines reference.[17]

Nevertheless, our data do not readily point to a particular explanation of the inconsistencies we found within participants with respect to their use of causal and descriptive information in making reference judgments, nor of the divergence in intuitions reported by Machery et al. Showing that a particular hybrid theory of reference is supported by folk intuitions would require uncovering the mechanisms governing the use of causal and descriptive information. Further research will be necessary to determine how participants are using each source of information, and in particular whether descriptive and causal information play distinct roles, as in Evans's theory, or are simply combined as independent sources of evidence. This last point is especially significant, because a hybrid theory that proposed that different individuals utilize both causal and descriptive information in a uniform way across all contexts would not predict the inconsistency that our experiments uncovered. Rather, our initial data support a theory with a more complicated account of the relation between the use of causal and descriptive information in different interpretive contexts. This suggests that views such as Evans's described above, which hold that that the causal source of a privileged set of descriptive contents associated with a referring term secures the reference of the term, are not supported by folk intuitions about reference. Given that participants were not consistent in judging either causal or descriptive information to be sufficient for securing reference, this suggests that the role that each plays is not fixed, as these views predict.

One possibility is that participants use descriptive information as a source of evidence concerning whether an individual obtained a concept from a particular causal source. For example, a layperson who believes gold has atomic number 79 may be judged more likely to be linked to the same baptism event as an expert metallurgist than someone who believes gold has atomic number 78, or that gold is a vegetable. On the other hand, causal information might be used to determine the significance of correct or incorrect descriptive information. Lack of appropriate causal grounding might suggest that possession of correct descriptive information is mere coincidence or lucky guessing. A further possibility concerns whether or not *individuating* descriptive information plays a privileged role in reference determination. Possession of key information about a concept, such as knowing the atomic number of a mineral, may make the presence of the false beliefs about the concept less relevant to judgments about concept possession. A complementary prediction is that radically false information, such as a belief that gold is a vegetable, may be sufficient to undermine other evidence for an appropriate causal history. If any of these possibilities is correct, it could be that participants vary in how heavily they weight different sources of information in inference, and not in the underlying referential practices the inferences inform.[18]

Another noteworthy consequence of our findings is that they provide preliminary empirical support for Burge's proposal that causal factors can at least partly explain

why individuals are sometimes willing to attribute a concept to someone with false beliefs about that concept. Although further research is necessary to determine more precisely the conditions under which individuals will make such attributions,[19] a hybrid theory of concepts seems to hold promise when it comes to solving the problem of ignorance and error.[20] In general, the mechanisms of deference and the ways in which the reliability of a source of information about a particular object or property affect judgments of reference require more investigation. Nevertheless, the fact that individuals are in some cases willing to attribute shared reference to someone using a concept who possesses false beliefs about it undermines a pure descriptive theory of reference, while the fact that an increasing number of false beliefs lowers the likelihood of attributions of shared reference shows that descriptive information is not completely irrelevant to such attributions.

Although our findings lend some preliminary support to a hybrid theory of reference and suggest some promising directions for future research, they are also subject to some important limitations. There are a number of methodological objections that have been raised against Machery et al.'s original studies and which might be applied to our experiments as well. To begin with, some philosophers have questioned the relevance of empirical data concerning non-philosophers' intuitions about semantics on both methodological and theoretical grounds. Devitt, for example, argues that the reliability of semantic intuitions depends on an individual's degree of expertise, so that philosophers' intuitions about reference trump folk intuitions: "just as the intuitions of paleontologists, physicists, and psychologists in their respective domains are likely to be better than those of the folk, so too the intuitions of the semanticists" (2011a, p. 7).

Devitt's claim makes sense for intuitions about the use of technical terms such as 'meaning' or 'reference', and it is also reasonable to suppose that specialists may be the ultimate authorities on the use of a particular name (such as 'Homer' or 'Shakespeare') or concept (such as ATOM or SPECIES). It is less clear, however, that expert intuitions are superior to folk intuitions when it comes to ordinary referential practices, or that there could be specialists in the practice of using names and concepts in general. It is an empirical question whether attributions of reference are sensitive to associated descriptive information (as the descriptive theory would predict) or to an individual's causal relations to the referent (as the causal theory would predict), and it is these ordinary folk attributions that influence everyday referential practices. Critically, it is not the use of the concept REFERENCE itself that has been studied in empirical work on semantics, but rather how individuals make reference judgments about particular uses of ordinary names and concepts. Assuming, as Devitt himself emphasizes, that the examples aren't too far removed from everyday cases of referential practice, folk judgments should be taken to provide evidence about the mechanisms underlying attributions of reference. So although we agree that it would be inappropriate to test folk intuitions about uses of terms like 'reference' or 'meaning', we see no difficulty with studying intuitions about uses and attributions of ordinary names and concepts by non-experts.

Nevertheless, other doubts about the relevance of folk intuitions about semantics remain. Several authors have criticized Machery et al.'s original experimental design as failing to distinguish different possible interpretations of their probes in ways that would explain the variation they discovered. Some (Deutsch, 2009; Ludwig, 2007) have argued that the experiments don't distinguish between semantic reference (what a name refers to) and speaker's reference (what the speaker refers to by using the name).[21] Sytsma and Livengood (2011) have argued that the wording of the reference questions posed by Machery et al. does not distinguish between the epistemic perspective of the narrator or reader (who knows that the descriptive information associated with the name is false) and the perspective of the character described in the probe.[22] Still others (Martí, 2009) have claimed that the experiments should test linguistic intuitions (intuitions about how a term should be used) rather than meta-linguistic intuitions (about what a term refers to), and proposed that experiments should test participants' beliefs about what a term refers to by asking factual questions about the referent of a term rather than questions about what a hypothetical person is referring to.[23]

These objections can be applied to our experiments as well. Our stimulus materials all focused around questions of whether two characters described in a vignette were "thinking about" the same natural or nominal kind. This formulation is open to the objection that the questions don't specify whether they should be evaluated from the perspective of the reader, who knows which of the character's beliefs in the vignettes are false, or that of the characters themselves. Furthermore, it is also open to the worry that there is an ambiguity between the semantic reference of the characters' concepts and the intended reference of the characters. Finally, one could worry that we solicited meta-linguistic intuitions, which are unreliable among non-experts, whereas we should have been probing linguistic intuitions.

We contend that our most surprising and important result, that of inconsistency within participants in their use of causal and descriptive information in making reference judgments, is not undermined by these methodological concerns. If participants did see our questions as ambiguous, we would predict resolution of these ambiguities in a uniform way within participants given that the questions asked and the context of the stories didn't vary between the relevant parts of the questionnaire. We would expect participants to evaluate the reference questions for either intended reference or semantic reference, and from either their own perspective or from the characters' perspectives, consistently across parts I and III. If this assumption is correct, then our finding of inconsistencies within participants with respect to their answers still suggests that they are using both causal and descriptive information in making judgments about reference rather than switching from one resolution of the ambiguities to the other between the two parts. If the context of the stories had varied in ways that led to systematically different assumptions about whether speaker or semantic reference was appropriate, or about the perspective from which to answer the reference question, this would have resulted in systematic differences between part I and part III judgments across participants. However, judgments across participants for part I and part III were remarkably similar in both experiments in *not*

reflecting a systematic approach to answering the reference question. Hence, the internal inconsistencies we found in participants' reference judgments indicate a genuine sensitivity to both causal and descriptive information, rather than systematic patterns of resolving questions about epistemic perspective or intended versus semantic reference.[24]

Nevertheless, the concern about soliciting semantic intuitions as opposed to meta-semantic intuitions suggests a solution to all three methodological objections that has so far not been attempted in the empirical literature on reference.[25] According to this methodology, the best way to probe participants' intuitions about reference would be to teach them novel proper names and other referring terms, and evaluate changes in their patterns of use of these terms as they receive new information about the terms. Participants could be taught a term, asked to use the term under conditions of limited and/or false information about it, and then prompted to use it further after being corrected. For example, participants could be taught that 'Racsagadam' is the name of an island off the coast of South America, as well as a number of facts about the island concerning its inhabitants, flora, and fauna. They could then be asked to make a number of inferences using the name based on what they have been taught about the island. Participants could then be told that the name originally referred to a section of the mainland, and asked to make new inferences using the name and to evaluate their previous answers. One factor that could be varied between participants would be whether or not the mistake about what the name applied to was limited to the participant herself or was widespread among other users of the name. Varying this and other conditions would make it possible to learn more about the roles played by causal and descriptive information, as well as contextual factors, while avoiding the ambiguities associated with soliciting third person judgments about reference.

The methodological considerations just discussed, along with the data we report above, suggest that the mechanisms underlying reference judgments may be complex and involve multiple variables. It is likely that most non-expert individuals lack explicit awareness of these variables, and that their intuitions depend on multiple layers of information and assumptions about the factors involved in reference in a particular scenario. What this suggests is that empirical work on reference should aim to tease apart the different theoretically significant factors influencing individuals' intuitions about reference, as opposed to focusing on determining whether such intuitions are consistent with antecedently formulated theories.

This leads to a final question about the relation between experimental work on semantics and Kripke's original arguments about the importance of causal connections for reference. Much of the recent discussion of intuitions about reference has centered around Kripke's Gödel example and others similar to it in structure. As several authors have pointed out (Devitt, 2011a; Ichikawa, Maitra, & Weatherson, 2011), intuitions about Gödel-style hypothetical cases were only a part of Kripke's overall argument against descriptive theories of reference. Machery et al.'s objective, however, is not to undermine a causal view of reference per se, but rather to cast doubt on the practice of supporting theories of reference on

the basis of intuitions about hypothetical cases (see also Machery, 2011; Mallon et al., 2009). Our view is that intuitions about hypothetical cases are not likely by themselves to settle debates over the nature of semantic reference, but rather are one factor among many—including a priori arguments such as those employed by Kripke and empirical findings about what factors drive semantic intuitions—all of which contribute to a better understanding of the nature of reference.

8. Conclusions and Further Research

To summarize, we have claimed that recent debates about semantic intuitions can be extended from the domain of proper names to that of concepts, given that the problem of ignorance and error facing theories of concepts parallels the difficulties for theories of proper names stemming from cases of false descriptive information being associated with names. We have also argued that these debates should consider the possibility of a hybrid account of reference in addition to pure causal and pure descriptive theories. Our empirical findings suggest that both causal and descriptive information play a role in individuals' judgments about reference, and as a result we believe that this framework should be evaluated in future empirical work on reference.

Along with many of the other authors mentioned above, we believe that future empirical studies need to avoid ambiguities that might lead participants to different interpretations of stimulus questions about reference, and some work has already been done in this regard (Machery, Olivola, & De Blanc, 2009; Sytsma & Livengood, 2011). Furthermore, we propose that along with posing factual questions about the reference of terms discussed in a vignette, experiments could be designed to teach participants novel concepts or names and track their patterns of use of these terms across a variety of conditions.

In addition to these methodological improvements, we want to suggest that future empirical research on reference should give more attention to investigating the potentially distinct roles played by causal and descriptive information in informing judgments about reference, rather than trying to solicit intuitions designed to support particular philosophical theories. Given the likelihood that the mechanisms governing reference judgments involve complex interactions between causal, descriptive, and contextual information, more work needs to be done to discover what each of these factors contributes to judgments about reference.

Acknowledgments

thanks also to Emily Wei and Paul Rogerson for help with experiments and data collection.

Notes

[1]   Some complications exist due to the fact that people can use the name 'Gödel' to refer to whatever they want (one can name one's pet 'Gödel,' for example), so in order to preserve the causal chain, a speaker must—at least when learning the name—intend to use it with the same reference as the person he or she learned it from. Whether such intentions and associated dispositions to defer to expert users of the name are part of sustaining the referential connection is controversial. For an argument against this view, see Devitt (2011b).

[2]   Such intentions wouldn't need to be explicit, of course, but could be demonstrated, as Burge suggests, by an individual's willingness to defer to those with a better understanding of the concept.

[3]   Expert users themselves will either stand in more immediate causal relations to what is picked out by the concept (much in the way one who is an "expert user" of a name will be acquainted with the referent of the name), or will stand in further appropriate causal relations to experts with immediate knowledge of whatever the concept picks out.

[4]   For further discussion of whether issues such as reference change motivate a move away from pure descriptive or pure casual theories of reference, see Devitt and Sterelny (1999, chapters 4 & 5) and Soames (2005).

[5]   In the manuscript posthumously published as *The varieties of reference*, Evans (1983) revised his view by claiming that the referent of a name is the causal source of descriptive information in the files of a privileged set of users of the name, and that for most users reference will be borrowed from the privileged users. Dickie (2011) defends a related view that improves on some significant deficiencies of Evans's later theory.

[6]   Machery et al. also prompted participants with a case in which the description associated with a name doesn't refer to anyone. In this condition, participants from both groups favored the causal theory, and did so at a higher rate than the Americans did in the Gödel case.

[7]   Machery (2009) cites earlier empirical work by Hewson (1994) as delivering a similar conclusion.

[8]   Few philosophers working on reference these days would endorse a pure causal or pure descriptive theory of reference. This would not be problematic for Machery et al.'s arguments if the versions of causal and descriptive theories that philosophers now defend still emphasized the opposition between causal and descriptive factors in determining reference. As we will suggest, however, hybrid theories of reference, which emphasize the importance of both factors, may be able to account for the variation in intuitions Machery and colleagues report.

[9]   More precisely, an ANOVA with the number of accurate responses as a dependent variable and domain (4: disease, mineral, artifact, legal document) and number of false beliefs (4: one, two, three, four) as independent variables revealed two significant effects: a main effect of the number of false beliefs, $F(3,176) = 11.75$, $p < 0.01$, and an interaction between the number of false beliefs and domain, $F(3, 176) = 3.03$, $p < 0.01$. True/false accuracy was highest when Alex held one or four false beliefs (one: $M = 23.00$, SD $= 1.03$; four: $M = 22.83$, SD $= 1.99$), and lower when he held two or three false beliefs (two: $M = 22.21$, SD $= 1.80$; three: $M = 21$, SD $= 2.61$). However, the only consistent pattern across individual domains was low performance with three false beliefs; the ordering for performance on other conditions varied. Given that the effect of number of false beliefs on accuracy was neither monotonic nor consistent across domains, it seems likely that

any differences in performance are unrelated to the variables of interest in the current experiments.

[10] Given our sample size of 192, this experiment had sufficient statistical power to detect a correlation as small as .18 with greater than 80% probability. Note that this correlation, which is between two binary variables, is technically a phi coefficient; however, computing the more familiar Pearson correlation generates the same numerical result.

[11] A Chi-squared goodness of fit test comparing the observed frequencies to those that would be expected from a normal distribution with the observed mean and standard deviation revealed that the observed distribution differed significantly from normal ($p < 0.01$).

[12] Because the distribution of ratings was not Gaussian, these data violate the assumptions of statistical tests such as the t-test and ANOVA. Key results were repeated with non-parametric tests, and revealed identical patterns of significance.

[13] Once again, a Chi-squared goodness of fit test comparing the observed frequencies to those that would be expected from a normal distribution with the observed mean and standard deviation revealed that the observed distribution differed significantly from normal ($p < 0.01$).

[14] Given our sample size of 128 in experiment 2, the experiment had sufficient statistical power to detect a correlation as small as .22 with greater than 80% probability.

[15] Recall also that the data Machery and his colleagues report (Machery et al., 2004, pp. B7–B8) does not show particularly high rates of agreement within cultural groups when it comes to privileging causal or descriptive information. This also suggests the possibility of finding empirical support for a hybrid theory of reference, since cultural differences in cognitive style do not explain within-culture variation in judgments.

[16] See Jackman (2009) for a related argument about Machery et al.'s conclusions.

[17] It is worth acknowledging that a pure causal theorist or a pure descriptive theorist could similarly claim that variation within and across cultures results from variation in additional assumptions that go beyond the information in the provided vignettes. For example, a causal theorist could claim that individuals vary in the inferences they make about a character's learning history, and descriptive theorists that individuals differ in whether or not they regard the provided descriptive information as constitutive of the concept or merely associated with it. This defense, however, is difficult to mount for Gödel cases like those used by Machery et al., where relevant facts about learning history and descriptive information are fairly explicit. Moreover, either approach would face difficulties in accounting for the novel data we report here, which includes effects of *both* common causal origins and shared descriptive information. To account for these effects, pure theorists of either variety would at minimum need to acknowledge that information of the other kind—shared descriptive information for a causal theorist and learning history for descriptivists—informs judgments, a possibility that is more readily accommodated by a hybrid view.

[18] For discussion of the significance of the centrality of descriptive information associated with referring terms to the referent's identity, see Devitt and Sterelny (1999, chapters 4 & 5) and Dickie (2011).

[19] This is particularly important given that in our experiments a greater number of false beliefs about a concept made individuals less inclined to give weight to shared causal history in attributing shared reference.

[20] In general, we believe that a hybrid theory holds promise for solving several of the more persistent problems facing theories of concepts, including the difficulty of formulating a theory that can address all of the main aims for which concepts are posited as an explanatory tool. For a dissenting view, see Machery and Seppälä (unpublished manuscript).

[21] For a reply to this objection, see Machery and Stich (forthcoming).

[22] The ambiguities pointed out by Deutsch (2009) and Ludwig (2007) and by Systsma and Livengood (2011) are not equivalent, since the participants have to decide whether to answer the reference question in terms of who is *actually* picked out by the descriptive information the character in the vignette associates with the name (the semantic reference as determined

by narrator/reader's perspective) or in terms of whom the character thinks that information picks out (the semantic reference from the character's perspective). This is distinct from the question of whether the character intends to refer to whomever is picked out by the descriptive information associated with the name, or the person whom other members of the linguistic community refer to using the name.

[23] Devitt (2011a) makes a related suggestion, claiming that it would be better to gather evidence for how reference works more directly than by studying intuitions about hypothetical cases. Devitt suggests creating vignettes in which the verbal behavior of the characters in vignettes involves uses of a name in ways that involve ordinary examples of ignorance and error (e.g., plausible false beliefs about an historical figure rather than fanciful counterfactual cases such as Kripke's Gödel case). He argues that prompting participants who are sufficiently knowledgeable about the referent of the name to attribute beliefs to the characters in the vignette would be a more reliable method of studying reference than asking them to make judgments about which person a use of a name picks out.

[24] A related shortcoming of our vignettes in both experiments is that we use rather than mention concepts in the various domains we tested in attributing beliefs to Alex in part I (as we do with the expert user of the concept). If participants were sensitive to this, they might interpret it as evidence that Alex possesses the concept in question. We find it very unlikely that participants interpreted the vignettes in this way, however, for if they did, we would expect much more systematic results in line with the causal theory for part 1, and a much lower effect of the number of his false beliefs. Moreover, if participants were interpreting the vignettes in this way, we would expect evidence of this to show up in their justifications of their responses to part 1 in experiment 2, but we found no such evidence in their justifications. The methodological suggestions we make in the next paragraph as to how to avoid problematic ambiguities in the vignettes would avoid this difficulty in future experiments. We are grateful to Michael Devitt for pointing this issue out to us.

[25] This methodology could be complemented by the approach suggested by Devitt (2011a), briefly described in footnote 23.

Appendix A: Stimuli for Experiment 1

Domain 1: Disease (tyleritis)

True beliefs (on Alpha):

1. Tyleritis affects the muscles and causes muscle pain.
2. Tyleritis is caused by exposure to a rare mineral.
3. Tyleritis can be diagnosed with a blood test.
4. Tyleritis can be cured by an injection.

False beliefs:

5. Tyleritis affects the joints and causes joint pain.
6. Tyleritis is caused by a virus.
7. Tyleritis can be diagnosed with a tissue biopsy.
8. Tyleritis is incurable.

Domain 2: Mineral (evensium)

True beliefs (on Alpha):

1. Evensium is a white mineral used to make white rings.

2. Evensium has a shiny finish.
3. Evensium is a good conductor of electricity.
4. Evensium is harder than gold.

False beliefs:

5. Evensium is a yellow mineral used to make yellow rings.
6. Evensium has a matte finish.
7. Evensium is a poor conductor of electricity.
8. Evensium is softer than gold.

Domain 3: Artifact (Krip-key)

True beliefs (on Alpha):

1. Krip-keys are magnetic and used to open locks.
2. Krip-keys were designed by the government.
3. Krip-keys can each be used to open only a single lock.
4. Krip-keys require the correct user's thumbprint to function.

False beliefs:

5. Krip-keys are digital and used to open locks.
6. Krip-keys were designed by private industry.
7. Krip-keys can each be used to open multiple locks.
8. Krip-keys require the correct user's DNA to function.

Domain 4: Legal document (Putnam)

True beliefs (on Alpha):

1. A putnam transfers the ownership of property.
2. A putnam must be signed by two witnesses.
3. A putnam cannot be revised after it is signed.
4. A putnam results in the transfer of ownership to the government in the event of the owner's death.

False beliefs:

5. A putnam transfers the ownership of a business.
6. A putnam must be signed by a judge.
7. A putnam can be revised up to a week after it is signed.
8. A putnam results in the transfer of ownership to nearest kin in the event of the owner's death.

References

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

Deutsch, M. (2009). Experimental philosophy and the theory of reference. *Mind & Language*, *24*, 445–466.

Devitt, M. (2011a). Experimental semantics. *Philosophy and Phenomenological Research*, *82*, 418–435.

Devitt, M. (2011b). Deference and the use theory. *Protosociology*, *27*, 196–211.

Devitt, M., & Sterelny, K. (1999). *Language and reality*. Cambridge, MA: MIT Press.

Dickie, I. (2011). How proper names refer. *Proceedings of the Aristotelian Society*, *111*. 43–78.

Evans, G. (1973). The causal theory of names. *Proceedings of the Aristotelian Society, Supplementary Volumes, 47*, 187–208.

Evans, G. (1983). *The varieties of reference*. Oxford: Oxford University Press.

Frege, G. (1948). On sense and reference (M. Black, Trans.). *The Philosophical Review, 57*, 207–230. (Original work published 1892.)

Hewson, C. (1994). Empirical evidence regarding the folk psychological concept of belief. *Proceedings of the Cognitive Science Society*, 403–406.

Ichikawa, J., Maitra, I., & Weatherson, B. (2011). In defense of a Kripkean dogma. *Philosophy and Phenomenological Research, 82*, 418–435.

Jackman, H. (2009). Semantic intuitions, conceptual analysis, and cross-cultural variation. *Philosophical Studies, 146*, 159–177.

Kripke, S. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press. (Original work published 1972.)

Laurence, S., & Margolis, E. (Eds.). (1999). *Concepts*. Cambridge, MA: MIT Press.

Ludwig, K. (2007). The epistemology of thought experiments: First-Person approach vs. third-person approach. *Midwest Studies in Philosophy, 31*, 128–159.

Machery, E. (2009). *Doing without concepts*. Oxford: Oxford University Press.

Machery, E. (2011). Thought experiments and philosophical knowledge. *Metaphilosophy, 42*, 191–214.

Machery, E., Mallon, R., Nichols, S, & Stich, S. (2004). Semantics, cross-cultural style. *Cognition, 92*, B1–B12.

Machery, E., Olivola, C., & De Blanc, M. (2009). Linguistic and metalinguistic intuitions in the philosophy of language. *Analysis, 69*, 689–694.

Machery, E., & Seppälä, S. (unpublished manuscript). Against hybrid theories of concepts.

Machery, E., & Stich, S. (forthcoming). Experimental philosophy of language. In G. Russell and D.G. Fara (Eds.) *Routledge companion to the philosophy of language*. Abingdon: Routledge.

Mallon, R., Machery, E., Nichols, S., & Stich, S. (2009). Against arguments from reference. *Philosophy and Phenomenological Research, 79*, 332–356.

Marti, G. (2009). Against semantic multi-culturalism. *Analysis, 69*, 42–48.

Putnam, H. (1975). The meaning of "meaning." In his *Mind, language, and reality: Philosophical papers* (pp. 215–271). New York: Cambridge University Press.

Searle, J. (1958). Proper names. *Mind, 67*, 166–173.

Soames, S. (2005). *Reference and description*. Princeton: Princeton University Press.

Sytsma, J., & Livengood, J. (2011). A new perspective concerning experiments on semanitic intuitions. *Australasian Journal of Philosophy, 89*, 315–332.