

Ambiguity and Referential Machinery

Ángel Pinillos

1. Introduction

One of the most important debates in the philosophy of language concerns the question of how proper names and natural kind terms manage to refer to things in the world. According to the descriptivist account, championed by Bertrand Russell and Gottlob Frege, the referent of a name or natural kind word N is just whatever satisfies some description competent users associate with N and believe about the referent of N.¹ In contrast, the causal-historical view focuses on the causal history of uses of the word. According to a very simple version of this position, championed by Saul Kripke, N refers to some object only if uses of N lead back, via a chain of communication, to a dubbing or some appropriate grounding of the term to its referent.² Crucially, there is no requirement that speakers have a description in mind that must be satisfied by the referent and that determines the referent of the term. The debate between the two camps has been intense, but in my view, it has obscured the possibility that both sides could be right. In particular, I am attracted to the thesis that names and natural kind terms are ambiguous between causal-historical and descriptive interpretations. An argument for the ambiguity of natural kind terms has already been put forward by Philip Kitcher (1978, 1983) and Nichols, Pinillos and Mallon (forthcoming). In this paper I raise some further considerations for the thesis and sketch how it might be extended to proper names. Although I report on two original experiments which give partial support for the thesis applied to proper names, these studies are modest and extremely limited. The main purpose of this paper is to lay conceptual room for the thesis. Much

more work needs to be done, but I think the possibility should be taken seriously.

2. Lewisian cluster view

Let me start by discussing the thesis as it applies to natural kind terms, using ‘water’ as an example. According to classic descriptivism, the term ‘water’ refers to whatever substance fits the description competent agents associate with the term and believe of the substance. If we consider a time before modern chemistry, the description associated with the word might be something like ‘The liquid that is clear, drinkable, and runs in rivers, streams and oceans’. Any liquid that satisfies the description will count as water, even if it is not H₂O. For example, suppose there is a far away planet, ‘Twin Earth’, that is just like Earth except that a certain molecule, XYZ, replaces H₂O. This liquid, XYZ, counts as ‘water’ (on the descriptivist account). In contrast, according to the causal-historical account, ‘water’ will refer to the liquid that is causally connected to our uses of ‘water’. This would only be H₂O and not XYZ. Here’s a case then where the competing views make distinct predictions. The sentence ‘Some water is XYZ’ would be true on the descriptivist reading and false on the causal-historical view. In contrast, according to the ambiguity hypothesis, some uses of that sentence are true and some uses are false, depending on the disambiguation of ‘water’.

In order to clarify the thesis, I want to discuss David Lewis’ (1994) view on the matter where according to him, natural kind terms are associated with a cluster of descriptions.

Although he doesn’t defend an ambiguity view of natural kind terms, his view resembles the ambiguity view (as I construe it) in certain important respects:

When we hear that XYZ off on Twin Earth fits many of the conditions in the cluster we are in a state of indecision about whether it deserves the name ‘water’. When in a state of

semantic indecision, we are often glad to go either way, and accommodate our usage temporarily to the whims of our conversational partners . . . So if some philosopher, call him Schmutnam, invites us to join him in saying that the water on Twin Earth differs in chemical composition from the water here, we will happily follow his lead. And if another philosopher, Putnam (1975), invites us to say that the stuff on Twin Earth is not water – and hence that Twoscar does not believe that water falls from the clouds – we just as happily follow his lead. We should have followed Putnam’s lead only for the duration of that conversation, then lapsed back into our accommodating state of indecision. But, sad to say, we thought that instead of playing along with a whim, we were settling a question once and for all. And so we came away lastingly misled. (Lewis 1994, p. 424)

I want to focus on two important features of this passage. The first is that the interpretation of ‘water’ is undecided or unsettled between the causal-historical and descriptive reading. For Lewis, the indeterminacy (for a token of the word) arises when we are asked to consider a place like Twin Earth, which fits many but not all conditions in the cluster. (In my view, the indeterminacy is more prevalent, as we will see.) The second is that the **precisification** is subtle in the sense that we may **precisify** (say by following Putnam or Schmutnam) without being consciously aware that the word **needs precisification** or being consciously aware that one is **precisifying** (this is the situation for most philosophers according to Lewis).

I accept both of these ideas, though I disagree with Lewis about the semantic mechanism **that** explains the phenomenon. Lewis accepts a cluster theory:

I think that ‘water’ is a cluster concept. Among the conditions in the cluster are: it is liquid, it is colourless, it is odourless, it supports life . . . but . . . there is more to the cluster than that. Another condition in the cluster is: it is a natural kind. Another condition

is indexical: it is abundant hereabouts. Another is metalinguistic: many call it ‘water’.

Another is both metalinguistic and indexical: *I* have heard of it under the name ‘water’.

(Lewis 1994, 424)

A cluster theory is not an ambiguity theory. I do not endorse Lewis’ theory for the following reason. There doesn’t seem to be much room for a causal-historical theory of reference. Saying that a term’s cluster contains a description with a relevant indexical (‘it is abundant hereabout’), or a description specifying that the denotation is a natural kind, is different from saying that a term’s cluster contains a description **that** must be satisfied by an object at the end of the causal-chain of communication. So I see no way that, in general, Lewis’ theory can accommodate certain uses where the causal-historical theory seems to get it right. Let us look at this in more detail.

The point can be made more vivid by focusing our attention on proper names. The descriptions most people associate with the biblical character Jonah are all mistaken, and yet, as Kripke (1980) pointed out, there are at least some uses of ‘Jonah’ that refer to a historical individual. If the best explanation of how the referent of *those* uses are fixed appeals to the causal-historical theory of reference, then this couldn’t be accommodated on the cluster view unless the cluster contained a description that mentioned the causal-historical chain (Ackerman 1979; Kroon 1987; Jackson 1998; Garcia Carpintero 2000). The problems with this view have been dealt with in detail by other philosophers (Devitt and Sterelny 1999; Soames 2002), and so I won’t go into details here. One problem is that it is not plausible that one of the descriptions typical competent speakers associate with Jonah is a description of the causal-chain mechanism (Devitt and Sterelny 1999). The causal-chain theory is at best an implicitly held theory and not specified by a name’s descriptive content. To see this, note that an important motivation behind

descriptivism is that the descriptions associated with a name provide the cognitive content of that name. As such, these descriptions often specify a person's mental content when that content is described using the name. But it is not plausible that ordinarily, these contents involve a specification of the causal-historical chain. For example, it doesn't seem like typical uses of 'Smith thinks Jonah lived in a whale' attribute to Smith a belief directly about a causal-historical chain of communication from uses of 'Jonah' tracing back to a dubbing or grounding on the historical figure. This belief is too theoretical for ordinary people to have. In addition, this account predicts that Smith's belief (that Jonah lived in a whale) is partially about the word 'Jonah'. But this is too strong a requirement since Smith might have forgotten Jonah's name. So Smith could still believe that Jonah lived in a whale, even if he failed to recall that the guy's name was 'Jonah'. This possibility would be ruled out by the view we are now considering.³

For these reasons, I do not believe that descriptive theories can accommodate many uses of proper names where the reference is fixed via a causal-historical chain. Of course, it is beyond the scope of this paper to fully argue that the descriptivist doesn't get it right in all cases. I refer the interested reader to Soames (2002) for detailed arguments against various descriptivist approaches. However, I do not conclude from this that the causal-historical account is fully correct. The ambiguity theory also vindicates the descriptivist account.

3. Indeterminacy and ambiguity

As I mentioned earlier, what I think is right, or approximately right, about Lewis' view is that there is indeterminacy with respect to the semantics of 'water', and that it can be made determinate by subtle changes in context. However, the indeterminacy in meaning for a token instance of 'water' is not generated just in conversations where the substance being described (XYZ) only partially satisfies the description cluster associated with the word (as Lewis seems to

suggest). Rather, in my view, the indeterminacy is the default setting for the word. This is very similar to how Lewis (1999) thinks about a different topic: vagueness. The word ‘house’ is vague. The extension is indeterminate between the set determined by the condition expressed by ‘is a main structure minus a garage’ and the set determined by the condition expressed by ‘is a main structure plus a garage’. Nothing in our practices settles that ‘house’ means one thing as opposed to another. In addition, there doesn’t seem to be great obstacle to adopting one or another interpretation depending on the interests and goals of conversational participants.

Lewis’ treatment of vagueness is naturally seen as a type of (massive) ambiguity. I think the same phenomenon is at play with proper names and natural kind terms. Moreover, both phenomena arguably have similar roots: the conventions of language fail to settle on one interpretation over the other. We may call the phenomenon concerning names and natural kind terms ‘vagueness’, but I prefer ‘ambiguity’. The reason is that, typically, when a word is vague, it often admits of multiple candidate extensions, but the phenomenon I have in mind restricts the interpretations to two: causal-historical and descriptivist.

The type of ambiguity at issue is not at all like that found for homonyms like ‘bank’ (between river bank and financial institution). With the latter type of ambiguity (but not for names, natural kind terms or Lewis’ ‘house’) the candidate meanings may be wholly unrelated. In addition, while we expect the disambiguation of ‘bank’ to have distinct lexicalizations across languages, we do not have the same expectation for names or natural kind terms, just as we do not have the same expectation for ‘house’.

It is important to clarify that the ambiguity view I am presenting is incomplete. I will not be providing the specification of the candidate meanings corresponding to the disambiguations of proper names and natural kind terms. Instead, I will just be arguing that these terms must have at

least two meanings corresponding to either the causal-historical or the descriptive reference fixing mechanism. Reference fixing mechanisms are not meanings, nor do they uniquely determine meanings. For example, a causal-historical account of names can be combined with a Millian semantics for proper names or a view under which the cognitive content of a name is a description.⁴ Of course, if we accept the latter, we will reject the idea that the referent of the name is just what satisfies the description associated with the term. The descriptivist account is also neutral with respect to a theory of meaning. Although the description which fixes the referent is normally taken to be or correspond to the meaning of the term (or a component of the meaning), there is no requirement that this be so.

So although the meaning of a term and its reference fixing mechanisms must be kept separate, the notions are intimately connected. If two reference fixing mechanisms predict different referents (or extensions) for a term, then they predict different meanings. For reference is certainly either a component of meaning or is determined by it. Hence, if we find that a term is ambiguous in reference or referential mechanisms, then there should be ambiguity in meaning. This is the strategy that I employ in this paper. I do not go further and specify what these meanings are.

4. The character of the data

Why do we think that 'bank' is ambiguous? Here's one important consideration. We can use the word to mean financial institution and we can also use it to mean riverside. At the very least, this type of consideration is sufficient to take the possibility of ambiguity seriously.

I submit that similar considerations can be raised in favor of the thesis that names and natural kind terms are ambiguous. I find David Lewis' suggestion that speakers could go either

way with the interpretation of ‘water’ extremely plausible. Another example is ‘fruit’. I think ‘fruit’ is ambiguous between a meaning that includes tomatoes and cucumbers in its extension and a meaning that excludes them. I also hold that ‘Jonah’, the name of the Biblical character whose generally associated descriptions are mostly false, can get either a descriptive or causal-historical interpretation. For example, if someone were to insist that Jonah wasn’t a real person on account of the descriptions associated with the term not being satisfied by a real person, then I would be happy to agree and go along with her. And if someone else were to insist that Jonah was a real person on account of the causal-historical chain leading to a real person, I would be happy to agree and go along with her as well.

Admittedly, examples of ambiguity for names and natural kind terms are more difficult to find and are less readily available than for ‘bank’. Why is that? The reason is that the distinct interpretations for names and natural kinds are normally in harmony.⁵ That is, normally, the distinct reference-fixing mechanisms yield the same extensions in the actual world (or at least in our immediate environment). The differences are subtle and often appear only when we focus on remote possible worlds or faraway places. With ordinary ambiguity, the distinct interpretations are normally well represented in our environment.

Of course, simple judgements about dual uses can only take us so far. There are a number of ways that the initial data may not signify ambiguity. There are alternative explanations that appeal instead to pragmatic effects, performance errors, context sensitivity or polysemy. I do not believe appealing to these ideas is promising. Or at least, they are less promising than the ambiguity view. I will revisit these ideas one by one after I present some initial data.

To end this section, I would like to make a remark about how the debate between descriptivists and causal-historical theorists is understood from the perspective of the ambiguity theory. A philosopher defending one of the two theories will often draw our attention to a sentence or linguistic construction **that** is then evaluated with respect to some possible scenario, and then the case is made that the resulting intuitive judgement favors one theory over the other. From the perspective of the ambiguity view, the conclusion is hasty. This datum does not favor one theory over the other. Instead, it merely indicates that in certain situations, a particular disambiguation is preferred.

5. Experimental data for natural kind terms

I want to begin my discussion with some recent work I have done in collaboration with Shaun Nichols and Ron Mallon.⁶ This research supports the ambiguity view of natural kind terms. One experiment we ran was designed so that exposure to or absence of a subtle cue about scientists misdescribing a kind would push subjects back and forth between causal-historical and descriptive interpretations of a term. We gave the following motivation for this experiment:

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 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. (Nichols et al. forthcoming, p. 11)

In fact, when Putnam argued against descriptivism for natural kind terms, he focused on hypothetical cases where a familiar kind was mischaracterized (gold, cats, etc.). From the perspective of the ambiguity theory, Putnam was merely pushing one interpretation. Our experiment exploits this insight.

In the neutral condition, we gave subjects this prompt:

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 his 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.

and in the experimental condition subjects got the following:

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 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.

Both groups were then given the following passage based on medieval bestiaries (Pliny the elder 1940 book 8 from Isidore of Seville 2006). One benefit of using this passage is that it involves an unfamiliar term, and hence it is less likely that subjects would import prior information about the kind to the experiment which can distort the results.

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.

Subjects were then asked whether they agreed with the sentence 'Catoblepas exist'. What we found was that subjects were swayed by the misdescription cue. They were more likely to say that Catoblepas existed when they were given the misdescription cue about Triceratops prior to the vignette. The ambiguity view can help explain this result since on the causal-historical reading, 'Catoblepas' should refer to wildebeests and on the descriptivist account, it should fail to refer (because the associated description is not satisfied by anything). The subtle cue about misdescriptions shift interpretations from one reading to another.

We explored this idea further in another experiment. This time, we deleted the priming material and just focused on the *Catoblepas* vignette (we modified it somewhat to make the causal-historical path to wildebeests clearer). We then asked participants to record their agreement or disagreement with the following statements (both in a ‘between’ and a ‘within subjects’ study):

- (1). *Catoblepas* exist
- (2). *Catoblepas* are wildebeests

What we found was that subjects rejected (1) (in accordance with the descriptivist account) but accepted (2) (in accordance with the causal-historical account). This is hard to explain on a univocal account of natural kind terms. This is because, on that view, if you are committed to (2), then you should also be committed to (1), precisely what was not found. This result is especially difficult to explain for the ‘within subjects’ condition where subjects saw (1) and (2) back to back. If there was an easy inconsistency to catch between them, our subjects should have detected it.

Although the ambiguity theory has the flexibility to account for the divergent responses to (1) and (2), there is still a need to give a more specific explanation about why subjects are more attracted to the causal-historical interpretation for (2) than for (1). In the [chapter](#), we argued that this result can be explained in part by appealing to presupposition accommodation.

Arguably, sentence (2) presupposes that *Catoblepas* exist. Therefore, subjects search for an interpretation of ‘*Catoblepas*’ [that](#) can accommodate this. The interpretation [that](#) achieves this is the causal-historical one linking the term to wildebeests. Crucially, the descriptive interpretation does not achieve this. In contrast, sentence (1) does not carry the presupposition of existence

(since this is what the sentence says). Subjects in this condition then are less likely to give a causal-historical response.

There are a couple of alternative explanations to the data, which can be pursued by univocal theorists. I want to discuss them here since they were not addressed in detail in the paper. First, it might be thought that (2) should be read as a universally quantified sentence ' $\forall x(Cx \supset Wx)$ ' where 'C' abbreviates 'Catoblepas' and 'W' 'Wilbeests'. This sentence is vacuously true in the cases in which the extension of 'C' is empty. Hence, this interpretation predicts the results we got. However, I do not think that the universally quantified sentence is a plausible interpretation of (1). That interpretation predicts subjects will judge that a replacement of the second predicate 'W' with any other would yield a judgement of truth. Thus, it would predict that 'Catoblepas are squirrels' and 'Catoblepas are triangles' are both true. Although we didn't experimentally test these cases, we find it implausible that subjects would agree with these sentences.

A second, more plausible alternative explanation can be raised by the defender of the univocal theory. It might be thought that our subjects surmise that catoblepas are a non-existing type or subspecies of wilbeest. Consider an analogy. Intuitively, we may agree with the claim that unicorns are four legged but reject that unicorns exist. Perhaps whatever explains this about unicorns can explain our data about catoblepas.

There are a number of things to say about this proposal. First, the view doesn't seem to be predicted by the causal historical view (at least a simple version of it). This is because if the causal chain of 'Catoblepas' really leads back to wilbeests (which exist), then Catoblepas should exist. Second, there is an important difference between 'Unicorns are four legged' and

‘Catoblepas are wildebeests’. The predicate in the latter can denote a natural kind while the predicate in the former **cannot**. This difference is important. Suppose we learn, for instance, that talk of unicorns was based on horses with branches stuck to their heads (just as talk of catoblepas was based on sightings of wildebeests). If we were tempted to say that unicorns were horses, we wouldn’t also be tempted to say that unicorns do not exist. Hence, it is not clear that the analogy with unicorns will be very helpful.

Third, the idea that catoblepas are a non-existing kind or subspecies of wildebeests could be interpreted in a variety of ways: (a) Catoblepas no longer exist but they used to (have gone extinct), (b) Catoblepas are a non-existing mythical creature and they are wildebeests, and (c) there is a myth according to which Catoblepas are wildebeests. Each of these interpretations is problematic and would saddle our subjects with serious confusion. We can rule out (a) since there is nothing in the vignette **that** suggests that Catoblepas used to roam earth but have now died off. We can rule out (b) because mythical creatures are not a type of mammal, so mythical creatures cannot be a type of wildebeest. Finally, (c) is problematic because it is not clear that even according to the myth, catoblepas belongs to an actual species (and, for instance, is a mammal). Rather, what seems plausible is that according to the myth, catoblepas are a magical animal and that discussion of catoblepas was inspired by wildebeests. Consider mermaids: **supposedly** talk of mermaids was based on sightings of sea lions. It wouldn’t be correct to say that in the mermaid myth, mermaids were sea lions. In the myth, they were hybrid fish-human creatures, not sea lions.

Of course, I cannot rule out the possibility that the subjects are confused, so it may be that our subjects accept (a), (b) or (c) (despite their implausibility). Absent some evidence that our subjects are going for (a-c), it is preferable not to attribute to them confused notions.

There are, then, a number of problems with the alternative explanations of why responses to (1) and (2) diverge. I think the ambiguity hypothesis gives the most plausible account.

6. Experimental data for proper names

Based on the previous experiment, I developed a vignette concerning a goblin named ‘Benjaminus’ with associated descriptions that aren’t satisfied by anything. However, the causal-historical chain does lead back to a real person named ‘Benjamin Smith’. So if descriptivism is right, ‘Benjaminus’ would fail to refer, but if the causal historical theory is right, then ‘Benjaminus’ would refer to Benjamin Smith. Here’s the vignette:

The medievals used to believe in a goblin named Benjaminus. Benjaminus was widely thought to fly around the night sky and also thought to have the power to read people’s minds. According to legend, Benjaminus liked to scare the local population by going in their houses and moving objects around. Of course, we know now that nothing in fact ever met the description associated with ‘Benjaminus’. However, historians in Germany have discovered that the tales of Benjaminus were based on a real person, Benjamin Smith, who was a skillful night thief in medieval London.

After the vignette, I asked participants to record their level of agreement with (a) ‘Benjaminus existed’, (b) ‘Benjamin Smith existed’ and either (c1) ‘Benjaminus is Benjamin Smith’ or (c2) ‘Benjamin Smith was Benjaminus’ (there were two conditions which differed only with respect to [c1]/[c2]). For each question, subjects marked their answer on a 5 point Likert scale (1=strongly disagree, 2=somewhat disagree, 3=neutral, 4=somewhat agree, 5=strongly agree).

I found no differences between the (c1) and (c2) responses, $t(35)=0.28$, $p>0.05$, so the following data collapses the two groups. In accordance with the descriptivist theory, subjects

tended to disagree with 'Benjaminus exists', $M=2.13$ ($3=\text{neutral}$) $SD=1.22$. This number was below the neutral point, $t(29)=3.90$ and the difference was statistically significant, $t(29)=3.90$, $p<0.01$.

Now, the large majority subjects tended to agree that Benjamin Smith existed. This information is explicit in the passage, so I excluded the eight subjects who failed to agree for failing to comprehend the passage. The mean for the identity statements 'Benjaminus was Benjamin Smith' / 'Benjamin Smith was Benjaminus' (recall there was no difference between these) was 3.41 , $N=30$, $SD=1.35$. And the difference between responses to the identity statements and the existence claim was statistically significant, $p<0.001$ $t(28)=5.38$ (paired t-test). The difference between responses to the identity statement and the neutral point (3) was not statistically significant, $t(29)=1.66$, $p>0.05$.

These results are difficult for the univocal theorist to explain. If people disagree that Benjaminus existed, and agree that Benjamin Smith existed, then they should disagree that they are one and the same. But this is not what was found. It was found instead that our subjects are neutral on this question. Moreover, as I mentioned above, the difference between these responses was statistically significant.

To see how puzzling these results really are, suppose that while people thought that Santa Claus didn't exist and that Bill Clinton existed, they were ambivalent about whether Santa Claus and Bill Clinton were one and the same. This type of result would be puzzling, though a natural response is to wonder whether subjects were using 'Santa Claus' (or 'Bill Clinton') in a uniform way across the questions. Similarly, I think that positing ambiguity can help explain the data.

The ambiguity theory has the resources to explain these results. When subjects consider (and disagree with) ‘Benjaminus exists’, they are using the descriptive disambiguation of the name, and when they consider ‘Benjaminus is Benjamin Smith/Benjamin Smith is Benjaminus’, they are more likely to invoke the causal-historical disambiguation.

So far so good, but why do subjects tend more toward the descriptive interpretation for the existence prompt than for the identity prompt? Perhaps the identity claim makes an interpretation of ‘Benjaminus’ that refers to Benjamin Smith salient. That interpretation is less salient when agents consider the existence claim. Hence, the identity prompt is more likely to elicit the causal-historical disambiguation of ‘Benjaminus’

There is an important disanalogy between the Benjaminus and the Catoblepas vignettes. In the Catoblepas vignettes, uses of ‘Catoblepas’ were perceptually grounded to sightings of wildebeests. In the Benjaminus case, uses of ‘Benjaminus’ were presumably not perceptually linked to Benjamin Smith. Instead, the passage suggests that the causal chain links back from uses of ‘Benjaminus’ to uses of ‘Benjamin Smith’ which are in turn perceptually linked to Benjamin Smith. As such, the sentence our subjects were asked about, ‘Benjaminus is Benjamin Smith’, involve two occurrences in the same causal chain of names. In sharp contrast, ‘Catoblepas are wildebeests’ do not involve occurrences in the same chain of communication. Rather, the occurrences belong to different chains but both grounded on wildebeests (On David Kaplan’s account of words, ‘Benjaminus is Benjamin Smith’ involves occurrences of the same name or word whereas ‘Catoblepas are wildebeests’ does not).⁷

It might be thought that this could make a difference to judgements about identity. To be precise, it might be thought that the strength of a judgement about ‘a=b’ where ‘a’ and ‘b’ are in

the same causal-historical chain will be different from judgements when they are not (leaving all else the same as much as possible). To test this hypothesis, I developed a vignette just like the one for Benjaminus except that instead of saying that stories of Benjaminus were based on Benjamin Smith, I said they were based on Karl Gustavson. The idea here is that it is unlikely that ‘Karl Gustavson’ would be seen as belonging to the same chain of communication as ‘Benjaminus’.

As reported earlier, the mean response for ‘Benjaminus was Benjamin Smith/Benjamin Smith was Benjaminus’ (recall also that which name goes first in the sentence made no statistically significant difference) was a 3.41 (where 1 is strongly disagree, 3 is neutral and 5 is strongly agree). The mean response to ‘Benjaminus was Karl Gustavson/Karl Gustavson was Benjaminus’ was 2.97, $SD=1.17$ $N=31$ (the order of the names did not matter, $p=0.45$ $t(29)=0.75$). Although this number is trending more toward the disagreement end of the spectrum than responses to the first set of identity claims, the difference between them is not statistically significant, $p=0.178$, $t(59)=1.36$. This means that we failed to confirm the hypothesis that identity statements with names in the same causal chain are more likely to be seen as true. However, I think further investigation in this direction might bear some fruit.

7. Alternative explanations

I have said that a reason to accept the ambiguity view is that we have an initial impression that a name or a natural kind term can be used to mean different things. In addition, I have provided some modest experimental evidence for the position. But this initial data may not hold up to further scrutiny. Here are some alternative explanations. I do not believe these options are compelling.

7.1 Speaker's reference

Sometimes we use a name with the intention to refer to someone who is not the conventional bearer of the name. If Bob is about to spill his beer at a crowded bar and I mistake him for Juan, I will warn him, 'Hey Juan, watch your beer'. In this situation, I intend to refer to Bob (speaker's reference) using 'Juan'. But the semantic referent of the name will still be Juan. Here's a type of case where we could get data about names having double uses but where we don't want to posit ambiguity. It might be thought that the data I presented earlier here can be explained by appealing to the distinction between speaker's reference and semantic reference.⁸

In response, note that the paradigm instances of speaker and semantic reference divergence involve the user of the name making a mistake about some feature of their environment. In the case we just saw, I confused Bob and Juan. But our experiments (Catoblepas and Benjaminus) do not have this feature. Our agents are given all the information about the subject matter (which is assumed to be accurate). Moreover, by design, the terms used are ones that they probably never encountered before, so it is unlikely that they would import prior confounding information about 'Benjaminus' or 'Catoblepas' to the case.

7.2 Conversational implicature

Sometimes we use sentences to convey propositions that go beyond what is semantically expressed by a sentence. Paul Grice (1975), through his theory of conversational implicatures, gave an explanation of how this can happen for rational and cooperative agents. Crucially, the explanation of how an implicature arises normally involves the apparent violation or flouting of a maxim by uttering S, which is resolved as being merely apparent by assuming the speaker is attempting to communicate something beyond the semantic content of S.

Grice laid out four maxims that he hoped could do the work in explaining how implicatures arise: Quality (speak truthfully and backed up by evidence), Quantity (be as informative as required but no more), Relation (be relevant) and Manner (be orderly, clear and avoid obfuscation). Here's an example of Quality at work. If someone says sarcastically to you, 'you look great' in a situation when you obviously look terrible, you would take your interlocutor to want to communicate something that is not the semantic content of the sentence uttered (that you look bad). Otherwise, the interlocutor would be violating Quality.

It is not obvious to me how appealing to implicatures can help explain the conflicting uses of names and natural kind terms we have seen in this chapter. To be specific, it is not clear what feature of the vignettes would lead subjects to interpret questions about the level of agreement with sentences such as 'Benjaminus exists' or 'Catoblepas are wildebeests' to be questions about other sentences or questions that are not about the semantic referents of 'Benjaminus' or 'Catoblepas'.

It might help to think about a case where there is a referential shift for names, and it is plausible that implicature is at play. I make use here of interesting work by Jennifer Saul (2007). Consider a use of the sentence 'Superman leaps more tall buildings than Clark Kent'. This may strike one as true even if one is aware that Clark Kent is Superman: if 'Superman' and 'Clark Kent' co-refer, then the sentence cannot be true under standard assumptions of compositionality. One way of accounting for the intuition that the sentence is true is that uses of that false sentence implicate a true proposition: that the person stage of Superman/Clark Kent associated with wearing a cape and tights leaps more buildings than the person stage associated with being a reporter and wearing glasses. The implicature is triggered, presumably, by flouting the maxim of quality (which requires, among other things, that speakers not say obviously false things).⁹

Something similar may be said about the existence prompts in our studies. Perhaps the causal-historical theory is correct and it is just obvious that Catoblepas and Benjaminus exist and so the question about existence is taken to implicate a question about whether the descriptions associated with those terms are satisfied.

One possible response to this alternative explanation is that in surveys participants are often asked obvious (and non-obvious) questions as checks to ensure they are not blindly responding to the material. Most subjects get these questions right, signifying that quality implicatures are not often triggered. In addition, in the Benjaminus prompts, I asked subjects about whether Benjamin Smith existed (which should be obvious given the vignettes), but the subjects overwhelmingly agreed that he did exist. If this question did not trigger an implication through a Quality violation, it is not likely that the question about whether Benjaminus exists *did either*. As far as I can tell, then, I do not see a way that appealing to conversational implicatures can explain the dual use data.

7.3 Context sensitivity and polysemy

Some terms shift their extensions from conversation to conversation. For example, ‘I’ will refer to different people depending on who is speaking. Ever since David Kaplan (1989), it has been common to say that paradigm instances of context sensitive terms have a stable meaning, a ‘character’, which can be modeled as a function from contexts of utterance to intensions. For example, the meaning of ‘I’ is a function *that* takes as input a conversational context and returns a constant function from worlds to the speaker of the context. In this sense, we can say that pronouns and other indexicals are context sensitive.

But not all terms with shifty extensions work like this. Polysemy does not work like paradigm cases of context sensitivity. A word is polysemous whenever it admits of different meanings that are nonetheless related. Sometimes the relatedness of the meanings is systematic in the sense that there are general rules that can transform one meaning to another. In other cases, there aren't such rules. An example of the former is 'cup', which can be used to mean a container or else a quantity ('I filled the cup with vinegar' and 'there are twenty cups of butter in that batch of cookies'). Evidence that the meanings are systematically connected is that we see other words that admit of similar ambiguity ('bag', 'box', etc.), and we also see that their translations to other languages preserve the polysemy. Unsystematic polysemy lacks these features. For instance, 'glasses' can be used to talk about a eyewear or a drinking containers (which are related), but the relation between these two meanings is not also present between other words and is not preserved under translation.

I think that context sensitivity is not the right way to understand the dual use of names or natural kind terms (at the very least they don't work like paradigm cases of context sensitivity). First, with many indexicals like 'I' and 'you' and demonstratives like 'this' and 'that' we can readily come up with an approximation for the meaning (character) associated with those terms ('refers to the speaker' for 'I', 'refers to the audience' for 'you', 'refers to a nearby object demonstrated' for 'this', 'refers to a distant object demonstrated' for 'that' etc.), but we draw a blank if asked to produce a meaning that would determine which disambiguation is invoked on a particular occasion for names and natural kind terms. To return to our old example, we wouldn't expect an interesting meaning for 'water' (that gives us the dual use) any more than we would expect one for Lewis' 'house'.¹⁰

Second, the various extensions of context sensitive terms are normally sets or objects that are disjoint or non-overlapping. For instance, the various referents of ‘I’ are distinct individuals and so on for ‘you’, ‘this’ etc. Of course, there can be overlap as when ‘this’ is used to refer to the leg of a chair and later to the chair itself. But this is the exception rather than the norm. In contrast, as mentioned earlier, the disambiguations for natural kind terms are often in harmony and will wholly overlap in the actual world as well as nearby worlds. Very often, we have to come up with far-fetched scenarios to tease apart descriptivist and causal-historical interpretations.

The phenomenon under investigation seems to have more in common with systematic polysemy. The candidate extensions of ‘water’ are related, and the ambiguity expected can also be expected for many words (all names and natural kind terms) and across languages. However, it fails to pattern with polysemy in that (as I emphasized earlier) the candidate extensions for names and natural kind terms are overlapping.

A worry about categorizing our phenomenon as polysemy is that we would also need to include Lewis’ ‘house’ and countless other vague terms in the category. Ultimately, lumping these terms alongside ordinary cases of polysemy may not be fruitful. But even if so, a subcategorization just capturing the dual uses of names and natural kind terms would have to be reified.

References

- Ackerman, D. (1979), ‘Proper names, propositional attitudes and non descriptive connotations’. *Philosophical Studies*, 35 (1), 55–69.
- Devitt, M. and Sterelny, K. (1999), *Language and Reality*. Oxford: Blackwell.

- Evans, G. (1973), 'The causal theory of names'. *Proceedings of the Aristotelian Society*, Supplementary Volume 47, 187–208. Reprinted in *Collected Papers* (Oxford: Oxford University Press, 1985), page references to reprinted version.
- Frege, G. (1893/1952), 'On sense and reference', in P. Geach and M. Black (eds), *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Blackwell.
- Garcia-Carpintero, M. (2000), 'A presuppositional account of reference fixing'. *Journal of Philosophy*, 97 (3), 109–147.
- Grice, H. P. (1975), 'Logic and conversation', in P. Cole and J. Morgan (eds), *Syntax and Semantics*, Vol. 3. New York: Academic Press.
- Isidore of Seville (2006), *The Etymologies of Isidore of Seville*. Trans. S. Barney, W. Lewis, J. Beach and O. Berghof. Cambridge: Cambridge University Press.
- Kaplan, D. (1989), 'Demonstratives', in D. Kaplan, J. Almog, J. Perry and H. Wettstein (eds), *Themes From Kaplan*, Oxford: Oxford University Press, pp. 481–563.
- Kaplan, D. (1990), 'Words'. *Proceedings of the Aristotelian Society*, Supplementary Volume 64, 93–119.
- Kitcher, P. (1978), 'Theories, theorists and theoretical change'. *The Philosophical Review*, 87, 519–547.
- Kitcher, P. (1993), *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford: Oxford University Press.
- Kripke, S. (1980), *Naming and Necessity*. Cambridge: Harvard University Press.
- Kroon, F. (1987), 'Causal descriptivism'. *Australasian Journal of Philosophy*, 65, 1–17.
- Lewis, D. (1994), 'Reduction of mind', in S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*. Oxford: Blackwell, pp. 412–431.
- Lewis, D. (1999), 'Many, but almost one', *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press, pp. 164–182.
- Nichols S., Pinillos A. and Mallon R. (forthcoming), 'Ambiguous reference'. *Mind*.
- Russell, B. (1919), 'Knowledge by acquaintance and knowledge by description', in *Mysticism and Logic*: London: George Allen and Unwin.
- Saul, J. (2007), *Simple Sentences, Substitution and Intuitions*. Oxford: Oxford University Press.
- Soames, S. (2002), *Beyond Rigidity: The Unfinished Semantic Agenda of 'Naming and Necessity'*. New York: Oxford University Press.

Notes

¹ Frege (1893/1952), Russell (1919).

² Kripke (1980).

³ Diana Ackerman (1979) defends a view where the description correlated with a name is sometimes a specification of the causal-historical chain. However, she does not endorse the idea that this description provides the cognitive content of the name or enters into the content of belief ascriptions, which may be the chief advantage of descriptivism. The view does not count as descriptivist on the criterion I proposed (which follows Kripke's)

⁴ Evans' (1973) hybrid view is causal-historical but posits a descriptive content.

⁵ I thank a referee for the 'Ambiguous Reference' paper for the term 'Harmony'.

⁶ Nichols, Pinillos and Mallon (forthcoming)

⁷ Kaplan (1990)

⁸ Kripke (1977).

⁹ Saul considers the implicature account, but she does not endorse it.

¹⁰ Even if a term is not context sensitive, it will still have a default character **that** will just be a constant function from contexts to the term's intension.