

In Preparation for

The Externalist Challenge. New Studies on Cognition and Intentionality, (ed.) Richard Schantz.

Reference, Causal Powers, Externalist Intuitions and Unicorns

Gabriel M. A. Segal

0. Introduction

In this chapter, I will compare and contrast singular concepts with what I call ‘nature concepts’, these being lay concepts of natural phenomena such as liquids (water) and types of animal (tiger). I will argue that the reference of singular concepts, which is obviously a variety of wide content, has a role to play in psychology. Singular concepts, I will claim, have both wide and narrow contents. In my view, no other concepts have wide contents. I shall give some reasons for supposing that psychology does not need to recognize any kind of reference-like content for nature concepts.

1. Wide Content and Causal Powers

I’ll begin by looking at an argument from Fodor (1987) that is meant to show that, in general, wide content should be kept out of psychology.

Fodor argued more or less as follows. "Categorization in science is characteristically taxonomy by causal powers. Identity of causal powers is identity of causal consequences across nomologically possible contexts." (Fodor 1987 44). Psychology is a science. It taxonomizes by content. Causal powers supervene on local microstructure. So, Fodor concluded, content supervenes on local microstructure.

If the argument were sound, it would follow that, in general, ‘characteristic’ scientific taxonomies would not distinguish micro-structural duplicates. But, it’s obvious that plenty of scientific taxonomy does distinguish micro-structural duplicates and takes relational properties into account. For example ‘island’, ‘crocodile’ and ‘planet’ all apply to an object partly in virtue of relational factors (being surrounded by water, having the right ancestry, being gravitationally bound to a star). Fodor was of course aware of this. He attempted to account for it by recognizing those relational taxonomies whereby the relational properties ‘affect causal powers’. Being a planet is thus taxonomic because ‘whether you are a planet affects your trajectory, and your trajectory determines what you can bump into; so whether you are a

planet affects your causal powers, which is all that individualism [i.e. internalism] asks for' (1987 43).

But that's not all internalism asks for. Internalism is about classifying micro-structural duplicates together. 'Planet' distinguishes micro-structural duplicates: a duplicate of, say, Jupiter that was not gravitationally bound to a star would not be a planet. An internalist about astronomy would demand an astronomical equivalent of narrow content. That would require there to be some astronomical property shared by all and only planets and twin planets. And I doubt that there is one, since just about any reasonably robust, non-self-illuminating object could be a planet, if it got into the right position vis-à-vis a star. I'd guess that the property (if any) shared by all and only such objects is not taxonomic in any science, let alone in astronomy. Let us, then, not be internalist about astronomy. And, if we like internalism (as I do) let us not rely on the soundness of any arguments using premises that cite perfectly general features of science.

So, something must be wrong with Fodor's argument. Either taxonomy in science is not characteristically done by causal powers or causal powers are not locally supervenient. I am not sure about the first disjunct. But the second one looks right.

Fodor's official view is that 'identity of causal powers is identity of causal consequences across nomologically possible contexts.' On the other hand, he also says that being a planet affects your causal powers. That would suggest that, say, Jupiter and Twin Jupiter have different causal powers in spite of being micro-structural duplicates. Jupiter, being a planet, can bump into things in ways that Twin Jupiter cannot. If we use 'causal powers' in the spirit of that thought, then identity of causal powers is partly context-dependent and causal powers aren't locally supervenient.

Suppose that Lara can bench-press 250 pounds on Earth. If she went to the moon, she could there bench-press 1,500 pounds. Would her causal powers thereby be changed? 'Yes and no' seems to be the best answer. We can distinguish two senses of 'causal power', one for a context-independent notion, the other for a context-dependent one. We might express the first by 'causal potentialities' and the second by 'active causal powers'. Lara has the same causal potentialities on Earth and on the Moon, but her active causal powers are greater on the Moon. Both notions are perfectly good. But one has to be clear which one is at work when discussing arguments about internalism.

There is nothing wrong with taxonomies that respect relational properties. In some cases, although not necessarily all, they work because they reflect active causal powers. 'Planet' is one example. 'Moving' is another, in statistical mechanics: a molecule in motion has different active causal powers from one at rest. It can help heat things up.

I would even argue with Fodor (1987) over what he took to be an uncontroversial example of a relational property that fails to affect causal powers (for reasons recognized by Fodor himself in his (1991)). Fodor defined 'is an H-particle' so that it is satisfied by a particle at t iff Fodor's dime is heads-up at t . Correspondingly a particle is a T-particle at t iff Fodor's dime is tails-up at t . Fodor thought it was obvious that he can't change the causal powers of all the particles in the universe just by flipping his

coin. He said (1987 p 34) "whether something is an H- (T-)particle is irrelevant to its causal powers. To put it a little more tensely, if an event *e* is caused by an H-particle *p*, then that same event *e* is also caused by *p* in the nearest nomologically possible world in which *p* is T rather than H". That's wrong. It all depends what kind of event *e* is. One could build an H-particle detector. The machine could work by detecting a particle, checking whether Fodor's dime was heads- or tails-up and classifying the particle accordingly. If the machine has classified *p* as an H-particle and Fodor turns his coin over, then the machine reclassifies *p* as T. Suppose that *e* is an event of an H-particle, *p*, causing the machine to classify it as H. Counterfactually, had *p* been T, then it would not have caused *e*. Evidently, just about any relational property can affect causal powers: there just needs to be a possible detector for it.

Of course, 'H-particle' and 'T-particle' are not scientifically useful terms. But that's not because the properties they pick out could not 'affect causal powers'. Rather, it's just because, as a matter of fact, nature doesn't have lots of H- and T-particle detectors, so there aren't any interesting generalisations about H- and T-particles that we might want to describe and understand.

So Fodor's argument fails. Perhaps, then, wide content has a place in psychology because it 'affects causal powers'. In the next sub-section, I shall give some reasons to think that it does, and say something about how.

1.1 Reference in psychology

I tentatively believe that there is one and only one kind of wide content and that it has a place in psychology: the reference of singular concepts, those concepts typically expressed by proper names and demonstratives. I'll illustrate the sort of role I have in mind by a story involving a proper name.

Lara is an archaeologist who comes across a papyrus apparently describing a ritual performed by an ancient Egyptian priestess called "Nefertoti". Lara becomes interested in Nefertoti. She pursues her research, and comes to believe that Nefertoti's remains might be found amongst certain ruins near Luxor, in Egypt. She makes her way there, and, after a thorough excavation, successfully locates a sarcophagus containing the remains.

One would suppose that the semantic relation of reference holding between the concept Lara expresses by "Nefertoti" (her "'Nefertoti' concept") and Nefertoti, has a role to play in explaining why Lara ended up successfully locating the remains. It would be very surprising if it were an accident that the remains Lara discovered just happened to be those of the referent of a concept that features in some of the beliefs and desires that led her to Egypt. But what kind of role is reference playing here? Actually, the story is a bit complicated. I will now try to tell some of it.

I take it that the supervenience base of the referential relation is partly intrinsic to Lara and partly relational. The intrinsic part presumably includes Lara's general competence with names, her (partly tacit) cognition of how the reference of a name is fixed, her knowledge of the particular term "Nefertoti" and maybe more. The relational aspect consists in a chain of linguistic events. At the start of the chain, a

practice of using "Nefertoti" with the intention of referring to a particular individual, A, is set up among some people who can identify A independently of their knowledge of the name. Gareth Evans (Evans 1982) calls these 'producers'. At some point, a producer uses the name in the company of someone who doesn't know A. This person, a 'consumer', acquires the name and uses it with the intention of referring to the person referred to by the producer. At some point, the name passes to the scribe who scripted Lara's papyrus, and thence to Lara herself.

The intrinsic and relational aspects of this supervenience base provide different aspects of the explanatory role of reference. The relational aspect is purely historical and so does not directly make any relevant difference to Lara's or her concept's causal potentialities or her active causal powers. To see this, compare the situation originally described to one in which everything is exactly the same, except in respect of the historical relations linking Lara to Nefertoti. Imagine that Nefertoti was a priestess called "Nefertoti" and that various documents featuring "Nefertoti" exist in Lara's time. However, bizarrely, the manuscript that Lara originally reads features an occurrence of the name "Nefertoti" that is unconnected with the practice of using "Nefertoti" to refer to Nefertoti. The manuscript is a fake, created by someone who had never heard of Nefertoti, and who came up with "Nefertoti" by coincidence. In such a case, Lara's causal powers would be just as they are in the original case. She still ends up locating Nefertoti's remains near Luxor.

Of course, in that bizarre situation, the initial desire she expresses by "I want to locate Nefertoti's remains" fails to get satisfied. But, even so, the desire has the same causal powers as the one that does get satisfied in the original case.

Moreover, the concept's having its reference relates in obvious ways to factors that do enhance its active causal powers. The historical chain exists as part of the name-using practice. And that practice left marks that are still present in Lara's era: in particular, all the written records that contain the name "Nefertoti". The existence of those records enhances Lara's and her concept's active causal powers by providing her with an intellectual path leading to knowledge of the whereabouts of Nefertoti's remains.

The internal aspect of the supervenience base figures in Lara's causal potentialities. The intrinsic properties of her cognitive endowment that allow her to understand "Nefertoti" are part of what explains her searching as she does.

Putting these aspects together, we could say that Lara's concept is associated with her having a 'cognitive grip' on Nefertoti. . Very roughly speaking, a concept *c* associates with a cognitive grip on *o* if and only if *c*'s owner is in a position to form a large number of specific beliefs involving *c*, and these beliefs constitute knowledge. The beliefs must be 'specific' in the sense that they concern properties exhibited by *o* in particular and not by the majority of other objects of *o*'s kind. The cognitive grip itself supervenes on two things: the intrinsic aspect of the supervenience base of reference and - not the relational aspect of the supervenience base, but - current elements of the name-using practice that those relational factors are part of.

Segal (forthcoming a) claims that singular concepts endow their subjects with a cognitive grip on their objects. But that is too strong. Suppose, for example, that just after Lara learns "Nefertoti", all traces of the name-using practice, except those in Lara's memory, were to disappear: the name is no longer written or spoken anywhere. Then Lara wouldn't have much of a cognitive grip on Nefertoti. But her concept would still have its reference.

I would tentatively suggest that the reference of a singular concept is often enough associated with a cognitive grip to make reference worthwhile for psychology. Consider an example from Peacocke (1992) of a subject, S, pointing to a man, M, that he (S) can see in a garden. If M had been in another position in the garden, then S would have pointed there instead. That sort of situation occurs frequently. S has a visually based singular concept referring to M and this fact enters into the explanation of why S points as he does.

The general structure of this second case seems parallel to that of the first one. Once again it is plausible that the referential relation supervenes on intrinsic and extrinsic factors. The extrinsic factors are historical (it's because it's M that caused the formation of the concept that the concept refers to M), hence do not directly affect S's current active causal powers in any relevant way. S's capacity to point to M is not directly affected by the fact that M, historically, had a causal role in the formation of the concept. (Suppose, for example, that another man, M2, had caused it (or one just like it) and was its referent, and there had been a quick switch whereby M took M2's place. S would still have the capacity to point to M.) But the historical factors are closely linked with present and future phenomena that do directly enhance S's causal powers. Quick switches don't happen very often and it's no accident that M is still standing there, reflecting light etc.. And those latter phenomena enhance S's active causal powers.

There is a class of singular concepts for which the set-up is importantly different: those whose reference is fixed by description. Both name-type and demonstrative-type concepts can work that way: "Let us call the inventor of the magnifying glass 'Zippy'", "A man defied the King last night. That man will suffer." In these cases, the semantics is still likely to be associated with a cognitive grip. But the relevant semantic property is quite different: it is reference-conditions, not reference. Please consider (1) and (2):

(1) *Nefertoti* refers to Nefertoti

(2) (x)(*Zippy* refers to x iff x alone invented the magnifying glass).

(1) describes a semantic relation that can only obtain if Nefertoti exists and if certain real relations obtain between her and possessors of the concept. (2) describes a semantic property of a "Zippy" concept that does not essentially involve any real relations to Zippy and does not require Zippy to exist.

If you have a non-empty "Zippy" concept then you will have some cognitive grip on the referent, just because you grasp the reference-fixing description. In these cases, the descriptive content does all the work, and the referent itself is idle. If the subject is prone to form and revise beliefs about the referent in a way that is sensitive to the referent's actual properties, this is not due to their having causal influence.

It's just because the subject is thinking about something as the possessor of those properties.

The manner in which reference matters to psychology, if the above ideas about cognitive grip are correct, seems to involve two related factors. First, if you have a concept that refers to something, then you are more likely to have desires relating to that something, than if you don't. Second, since reference is associated with cognitive grip, you have a non-negligible chance of satisfying those desires, since you are in a position to gather truths about the thing. Thus, in general, if you have a concept that refers to something, there is non-negligible chance that you will end up trying to relate to it in certain ways and that your endeavours will be successful. The chances might be non-negligible enough for it to be worth a psychologist's paying attention to the relation of reference.

If that is a correct account of reference in psychology, then it follows that reference can be a causally efficacious property of a concept. So if one says, for example, that it is in part because she has a concept that refers to Nefertiti that Lara located Nefertiti's remains, the "because" is a genuine causal "by cause of". Or so I will argue, following up on an idea from Segal and Sober (1991).

Segal and Sober offer a sufficient condition for a property to be causally efficacious with respect to a particular effect. Essentially, the idea is that a macro-property *F* of an event, *c*, is causally efficacious in respect of some effect of event *e*'s having property *G*, if (a) *F* events are (*ceteris paribus*) nomologically sufficient for *G* events and (b) *c*'s possession of *F* mereologically supervenes on some set, *M*, of micro-properties of *c* such that *c*'s possession of at least one of *M* is causally efficacious in respect of *e*'s being *G*.

Segal and Sober claim that according to their condition, representational content can be causally efficacious. Peacocke (1992) objects. He takes up Segal and Sober's example of the presence of air being efficacious in respect of the lighting of a match. The criterion legitimates the air's claim to be efficacious even though it is the presence of oxygen that is doing the causal work, not the presence of the nitrogen that is also in the air. Peacocke says that according to Segal and Sober's treatment, semantic properties of a mental representation are no more efficacious in respect of the effects of that representation than is the presence of the nitrogen in the air efficacious in respect of the match's lighting.

Peacocke offers no argument for this claim and it appears to be false. If you apply Segal and Sober's condition to the atmosphere's property of containing nitrogen, then it does not license the efficacy of that property. Now suppose that we apply the condition to the reference of a singular concept. And let's suppose it's a case where the supervenience base of reference has intrinsic and extrinsic components, as discussed. Then reference meets at least one relevant requirement: an aspect of the supervenience base, the internal one, is itself causally efficacious in respect of relevant effects. And that seems right. Reference is not analogous to the presence of nitrogen in the atmosphere, but to the presence of air. It is the historical component of the supervenience base of reference that is analogous to the presence of nitrogen.

It is actually the other requirement, *ceteris paribus* nomological sufficiency, that appears more

problematic. Consider the pointing case. It seems true that, *ceteris paribus*, anyone who has a visually based demonstrative concept, referring to an object *o*, who tries to point to *o*, will point in *o*'s direction. But in Lara's case it's not so clear. Would anyone who has a name-concept referring to *o* and who wants to find *o* (or what's left of her) be likely to find *o*? That's not obvious. If what reference is doing here is enhancing people's chances of success, then the *ceteris paribus* nomological sufficiency requirement might be asking too much of it. It's not plausible that if we have a concept of something then, *ceteris paribus*, we'll be successful in our endeavours to relate to it. Personally, I find that my endeavours are often in vain.

Rather, it has more to do with non-negligible chances, as discussed above. Thus someone who has a concept referring to *x* and tries to relate to *x* in way *w* is more likely to relate in *w* to *x* than someone who doesn't have a concept referring to *x*. But, anyway, it's a fine line between 'ceteris paribus' generalisations and increased probabilities. There's no huge difference between, 'Ceteris paribus, if *p* then *q*' and '*p* increases the probability of *q*'. A version of Segal and Sober's condition that replaced the nomological sufficiency requirement with one framed in terms of raised probabilities looks as plausible as the original. And referential relations can meet this new condition for causal efficacy.

Before moving on to the contrast between singular concepts and nature concepts, I pause to note that all of the above is consistent with reasonable versions of Fregeanism and internalism. Thus, consistently with the above suggestions about reference, we may suppose that singular concepts have, in addition to reference, some other kind of content resembling Fregean sense, in some or other important respects. We may also suppose that, in addition to reference, they possess narrow content. And, in an Ockhamist spirit, we might go so far as to suppose that sense-like content is narrow and consists in, or determines, extension conditions.

In the next section I will argue for a contrast between non-empty singular concepts and concepts of other sorts.

2. Nature Concepts, Externalist Intuitions and Unicorns

The reference of certain singular concepts, then, may be a relation that matters to psychology. However, I would argue, reference is not the only kind of content required by psychology. Reference is a real relation, a relation that is only instantiated when both relata exist. On the face of it, there are plenty of singular concepts that don't refer but that still have cognitive content. For example, suppose that an ancient Egyptian, Ahmenphut, prayed to Ra to end a heat wave, because he believed that Ra was the god who was responsible for the Sun, and hence the god with most control over the temperature. Ahmenphut's "Ra" concept has a cognitive content that it contributes to the beliefs that motivate Ahmenphut's action. Evidently this content is not its reference. So we must recognize cognitive content that isn't reference.

On the face of it, the same reasoning appears to apply to a very wide range of non-singular concepts: *vampire*, for example, has an empty extension. And while concepts like *tax*, *taxon*, *tomfoolery* and

tomorrow, have non-empty extensions, it does not seem plausible that their having those extensions has any relevance to their cognitive content. That is to say, the concepts could have had the same cognitive content, had they had different or empty extensions.

If that's right, then it looks as though non-empty singular concepts may be rather exceptional. Perhaps they are the only ones whose actual extensional relations are of interest to psychology. In fact, I think they are, although I will not argue the point here.

The diametrically opposed view would, I suppose, be along the following lines. Reference of non-empty singular concepts is the paradigm of content, and we should seek an account of psychological content that is modelled on that paradigm. Empty singular concepts should be regarded as aberrations of some kind, and left out of consideration in theories of content. When it comes to general concepts, we need to find some analogue of reference to serve as content. That is, we need some real relation between concept and something else (say, a property or kind or substance) that can do some work in psychology analogous to that done by singular reference.

An intermediate view would be that there are some types of non-singular concepts that enter into semantic relations with real things in the world in a way that is psychologically significant and other types that don't. The obvious candidates for non-singular analogues of singular concepts are nature concepts. A popular view of the functioning of these concepts, arising from work of Putnam (1975) and Kripke (1972) sees nature concepts as very much like name concepts.

I shall state a case against the intermediate view. If that case is successful, it should cast some doubt on the extreme view. If we can't even find a role for the reference of nature concepts in psychology, then what hope is there for the others?

I will proceed as follows. First, I will briefly sketch the Putnam- Kripke-inspired account. Then I will briefly state some of the chief arguments against the view that have been developed elsewhere by others and myself. I will then focus on what I take to be one important source of support for the view: externalist intuitions about the semantics of nature terms. I'll argue that there are specific reasons why those particular intuitions should not be given much credence.

2.1 The alleged analogy between name- and nature concepts and a tempting thought

A nature concept is a lay concept of some natural phenomenon such as water, aluminium, polio, lemon or tiger. I begin by recalling the chief points of the Putnam- and Kripke- inspired account of nature concepts, staying closer to Putnam's version.

Be so kind as to envisage a lay subject, Mathilda, in 1250, and her understanding of the term "nightingale". She is reasonably good identifying nightingales, knows what a typical nightingale looks like and knows something about typical nightingale habits. She uses the term "nightingale" with the intention of referring to a particular species of bird, members of which she and others on Earth have

interacted with.

On a twin Earth, there are birds that look just like nightingales, but they share no ancestry with Earth nightingales and have different genes. Twin nightingales are not nightingales, they just look like them. If Mathilda were to encounter a twin nightingale and say "Lo, a nightingale!", she would be saying something false in her language. That is because she takes her term to be true of members of the nightingale species, not of anything that happens to resemble such birds. When she says "nightingale", she means *nightingale*, just as we modern sophisticates do. Mathilda does not know what makes a nightingale a nightingale. But she regards the species as having a hidden real essence. This essence is something that could be — and indeed later will be — discovered by science.

So Mathilda uses "nightingale" as if it had been introduced to the language by a dubbing ceremony: "We shall call those birds 'nightingales', and the term is also to be true of all and only those birds that belong to the same species of bird as do they." The real essence, which determines the sameness relation, is not known to the dubbers. But there is a fact of the matter about it all the same, and it might be discovered by empirical investigation.

It is tempting to think that "nightingale" works in much the same way as "Nefertiti". Speakers have a "nightingale" concept the extension of which is essentially related to a single species. This concept functions in such a way that it associates with a cognitive grip on the species. Thinkers deploy the concept in thought in such a manner that they form and revise beliefs involving that concept in way that is sensitive to the actual properties of the species.

We can make the suggested parallel with singular concepts more explicit: if nightingales, qua nightingales have feature F, then there is a reasonable chance that someone with a concept that essentially extends over the nightingale species will, if the F-ness of nightingales matters to her, come to believe that nightingales are F and will therefore have a fair chance of interacting successfully with particular nightingales.

If that suggestion is right, then we can see one reason why it makes sense to see nature concepts as having their extensions essentially limited to natural kinds rather than motleys. For natural kinds are the kinds that do best in inductive inferences. Nightingales and twin nightingales might look alike. But that's just a superficial similarity. Since they are different species, the underlying difference is more important than the superficial similarity. In the longer term, the differences between nightingales and twin nightingales will lead to cognitive differences between Earth and twin Earth thinkers: the former will come to know about and adapt their behaviour to nightingales, the latter will come to know different things about the other sort of bird, and will modify their behaviour appropriately. Hence it makes sense to assign difference content to Mathilda's and Twin Mathilda's "nightingale" concepts.

The thought that nature concepts work like that might be tempting. But I think it is wrong. Outside science, we simply do not have much in the way of cognitive grips on most of the natural kinds whose samples we bump into. The result is that we would get little predictive mileage by assuming that nature

concepts have their extensions essentially tied to natural kinds.

The analogy with name concepts fails:

(1) We are good at keeping track of individuals whose name we know: we are good at re-identifying them (finding out about their properties at different times) and we

don't often confuse them with other individuals. By contrast, before the relevant science is done we are often bad at re-identifying natural kinds. When Archimedes figured out how to distinguish gold from fool's gold, he made history. In the normal run of things, for most kinds of natural kinds, we have little idea which individuals belong to the same natural kind and which don't.

(2) We are very interested in keeping track of individuals. Our interests in kinds are many and various and typically not best served by carving nature at its joints.

(3) There are actual dubbing ceremonies for individuals. Outside of science, there are no dubbing ceremonies for kinds.

(4) We know what kind of thing a name is supposed to refer to: an individual.

But even if there were dubbing ceremonies for kinds, we would have no good idea of what kind of kind we would be referring to — subspecies, species, genus, family?

Further, as has often been pointed out, the Kripke-Putnam model makes a false prediction about actual usage. Suppose that, pre-scientifically, people apply a term to samples nearly all of which belong to the same natural kind, but occasionally fail to apply it to samples of the kind and/or apply it to some things that are not of the kind. Then the science is done and the facts discovered. The model predicts that, in such cases, we would revise our usage accordingly and correct what would seem to have been classificatory errors. In fact, though, this is not what typically happens. Rather we just carry on as before. "Cat", in its most general sense, is a good example of this. It applies to members of the family Felidae, but also to a few non-members, such as civet cats.

In the final sub-sections, I will suggest an alternative account of nature concepts, and go on to say something about the role that intuitions ought to play in adjudicating between the two accounts.

2.2 A motley alternative

The Kripke-Putnam account, then, is highly problematic. Let me offer an alternative. I suggest that nature concepts, rather than extending over natural kinds, extend over motleys. I expect that "motley" is

itself something of a motley term. But the idea is that what groups samples together under a motley concept is not a hidden real essence. Rather, members might be grouped either by features explicitly known to those whose concept it is, or by features that are determined in more (or less) complicated ways by features of the thinkers' cognition. The concept of tree as expressed by the Oxford English Dictionary, definition b., is a good example: "a. a perennial plant with a woody self-supporting main stem or trunk when mature and usu. unbranched for some distance above the ground. b. any similar plant having a tall erect usu. single stem e.g. palm tree".

The key point about motley concepts, when it comes to the issue of the role of reference in psychology, is that they act like descriptive concepts. It is their extension conditions, not their actual extensions, that matter to psychology. One can have a motley concept whether or not the motley exists, and the extension conditions of the concept are independent of the nature of any samples that there might be. Any cognitive grip on the motley that associates with possession of a motley concept supervenes on features of the thinker's cognition that are not essentially dependent on real relations between thinker and members of the motley.

And a key point about motleys when it comes to the debate over internalism is that it's plausible that motley concepts are shared by twins. Mathilda's and Twin Mathilda's "tree" concepts have the same extension conditions. If their "nightingale" concepts are motley concepts, then it's reasonable to suppose that they would both extend over both nightingales and twin nightingales.

2.3 Unicorns, externalist intuitions and chachalacas

One key source of support for the Kripke-Putnam account comes from certain externalist intuitions, for example, intuitions generated by classical twin Earth thought experiments. So, for instance, we are supposed to have the intuition that if Mathilda were confronted with a twin nightingale and said "Lo, a nightingale!", she would be saying something false and expressing a false belief. I think that there are specific reasons why we should not trust those intuitions, which I will explain as we proceed.

To begin with, contrast those intuitions with analogous ones about name concepts. Suppose that Mathilda's best friend is Mary, and that on Twin Earth, Mary has an identical twin. If Mathilda were to come across Twin Mary and say "That's Mary" then she would be saying something false and expressing a false belief. Now that is intuitively obvious. But the key thing is that it is obvious to just about everybody, including Mathilda herself. Our intuitions about our own use of names and the functioning of our own name concepts provide evidence about their meanings and reference.

By contrast, twin Earth intuitions relating to nature concepts are notoriously unstable, varying with how they are elicited and whom you ask. Does your intuition really tell you that Mathilda's concept isn't true

of twin nightingales, that it doesn't extend over the motley, rather than the species? Wouldn't we do better to ask Mathilda?

I suspect that both Putnam and Kripke draw on their intuitions about the extension conditions of concepts and mistakenly think that their intuitions are 'ours', that they are representative of those of all sensible, reflective humans. I think that there is a dangerous element of a priorism in that approach to the individuation of contents. The remainder of the paper will elaborate on that thought.

I begin with some claims made by Kripke in his unpublished (1973). In the course of an argument for the conclusion that unicorns could not have existed, Kripke made some remarks about the medieval use of the term "unicorn". Kripke introduced the possible term "shunicorn", which he defined as "member of the species of animals, if there in fact is one, which is identified by the following surface characteristics, having one horn and looking like a horse". He went on to claim that he doubted that this was the way the term "unicorn" was used by the medievals, who took the term seriously. He said that they probably did believe that there was only one species that looked like this. But they would acknowledge the possibility of species that had the relevant surface characteristics but that would not be unicorns. Kripke asked his audience to suppose that the medievals believed that Sir Galahad had met a unicorn. Then, according to Kripke, a medieval would agree that, if in some part of the world there was a species of animals other than that of which a specimen was met by Sir Galahad, but which looked like it, these animals would not be unicorns.

The thing to note is that Kripke was making substantial empirical claims about how the medievals used or would have used the term "unicorn". But Kripke provided no empirical evidence in favour of these claims. I would speculate that he thought empirical research wasn't necessary because he assumed that his own intuitions about species and our words for them would be a reliable guide to medieval intuitions about unicorns and "unicorns". But given the huge intellectual differences between 20th century academics and medieval lay persons, such an assumption would be unwarranted.

I have done a little of the required research and it does not confirm Kripke's claims. First, it is widely accepted among medieval historians that the term "unicorn" was used by the medievals as a broad motley term meaning *quadruped with one horn*. So, for example, there are two species of unicorn described in Androvandus' 12th century Bestiary (from the translation, White 1954). Here are extracts:

"Unicornis the Unicorn, also called rhinoceros by the Greeks, is of the following nature. He is a very small animal like a kid, excessively swift, with one horn in the middle of his forehead and no hunter can catch him." (p. 20) That sort of unicorn can only be caught by a lone virgin in the woods. And the second extract tells us: "The monocerous is a monster with a horrible howl, with a horse-like body, with feet like an elephant and with a tail like a stag's. A horn sticks out from the middle of its forehead with astonishing splendour" (p. 44). Unicorns of that sort cannot be caught at all, although they can be killed.

The term "unicorn" was also used more specifically to designate a specific type of unicorn. In that usage, would it have expressed a motley concept or a natural kind one? Kripke's claim was that a medieval

would agree that if there were animals that looked like members of the species a specimen of which was encountered by Sir Galahad, but belonged to a different species, then they would not be unicorns. Perhaps in some sense Kripke is right about that. But if so, then it would not show that the medieval species concept of unicorn was a natural-kind concept rather than a motley concept. There are two related reasons for this, one to do with Kripke's implicit choice of question and one to do with the term "species".

Kripke's implicit question is framed in terms of appearance: would these animals that looked like unicorns be unicorns? But motley concepts are not typically appearance concepts: being an F might involve a great deal more (and/or less) than looking like an F, even if F-hood is not a natural-kind property. And the problem with "species" is that we are used to using the term in its contemporary scientific sense. The medievals did not use it that way. It is therefore not particularly clear how a medieval would have interpreted a question framed in terms of "species", nor is it clear that we could have asked them about conspecificity without first teaching them biology.

I suggest we consider a variation on Kripke's thought experiment. Suppose then, we were to ask Mathilda this:

"In a land far away, there are some animals that resemble the unicorn encountered by Sir Galahad. This type of animal has one horn in the middle of its forehead, it ventures to attack elephants, so sharp is the nail in its foot that with one blow it can rip the belly of a beast. Hunters can catch the animal only by placing a young virgin in its haunts. No sooner does the animal see the virgin, than he runs towards her and lies down at her feet and so suffers himself to be captured by the hunters. But these animals are of a different species from the one encountered by Sir Galahad. Would they be unicorns?"

How might Mathilda have reacted? I think she might have been rather mystified. She might wonder how an animal with so many of the characteristic properties of unicorns could be of a different species. In virtue of what would they not be unicorns? Thus she might have said: "The animal you have described is certainly a unicorn. Your description is most precise. Indeed, the unicorn is the only animal so brave as to venture to kill an elephant. Please explain in what sense they belong to a different species from the one encountered by Sir Galahad". An answer in terms of shared ancestors or genes or interbreeding would not have meant much to Mathilda, since the medievals did not think of species in those terms.

In sum, Kripke's claims about what a medieval would agree to are based on his own intuitions. And one would suspect that those intuitions are partly based on his reading key features of his own contemporary, scientifically-informed concept into the pre-scientific medieval concept. Kripke provides no empirical evidence in support of his claim, and certain evidence suggests that it is false. I move on now to some more evidence bearing on the nature of nature concepts. .

Anthropologist Scott Atran and his collaborators made an extensive ten-year (and counting) study of what they call "folkbiology", this being the way normal humans think about plants and animals. Atran studied, in particular, people of various tribes with little or no scientific education, living in natural

habitats around the world. Atran and his collaborators found interesting universals. In general, such people are very good at classifying (middle-sized and large) flora and fauna. Folk classifications for these correspond closely to scientific ones in respect of their actual extensions. The folk classification scheme is always organised as a ranked taxonomy, that is, a genuine hierarchical taxonomy that allows for generalisations over the different ranks themselves. And, finally, the folk believe in individual essences. Each individual plant or animal has an essence that determines its identity and persistence conditions. (See e.g. Atran 1999). Pinker (1994 423) recounts: in rural Nigeria, some students disguised a pawpaw as a pineapple and asked natives whether the result was pawpaw or pineapple. A typical response was 'It's a pawpaw because a pawpaw has its own structure from heaven and a pineapple its own origin. One cannot turn into the other.' Such essences are species-specific and passed from parent to offspring.

Interestingly, these apparent universals of folkbiology are not apparent in children who grow up in cities and who are explicitly taught biology (see Carey 1985). But city dwellers are exceptional in their biological ignorance. Atran's work indicates that there is an innate, species-specific folkbiology module of some kind, which module functions in humans who grow up in our natural environment but which atrophies in humans who grow up in cities and receive formal education

Kripke and Putnam tried to find out about human nature concepts, including specifically those of plants and animals, by consulting their own intuitions. But they, too, grew up in cities and received formal education. This means that their intuitions may be far from representative.

In particular (as Atran thinks) the view that natural kinds are individuated by hidden, real essences may be the fruit of an illegitimate union between the folk idea of heaven-sent essences and a respect for science as the discoverer of deep truths. But *heaven-sent* essences are not the kinds of things that science can tell us about. They are other-worldly.

Putnam's and Kripke's intuitions are suspect. The intuitions of the folk themselves, however, can provide useful evidence about the nature of the folks' own concepts. We can't ask Mathilda about her intuitions. But we can ask the folk. By consulting people from unscientific cultures living in nature, we have a chance to study nature concepts in something like their natural habitat. Such a study might provide evidence relevant to adjudicating between the natural-kind hypothesis and the motley hypothesis.

Atran, his collaborators and I have made start. Atran and I constructed a questionnaire designed to tap relevant twin-Earth intuitions among tribespeople. The questions were put to Maya subjects in the Yucatan, Mexico. Unfortunately, at the time of writing some of the data remain inaccessible, due to floods and hurricanes in Mexico. As a result, we have not properly analysed such data as we have. I conclude with an anecdote arising from our pilot study.

The key question that we asked was as follows:

"In a far away land there are some birds. These birds look exactly like the

chachalacas here in Yucatan and move and sound exactly like the chachalacas here. But these birds do not come from the same ancestors or lineage as the chachalacas here. Are the birds of the far away land chachalacas or not?"

What answers might a proponent of the Putnam-Kripke account predict?

Perhaps one sensible answer, under this view, would be "maybe". The question would be whether the twin chachalacas had the same hidden, real essence as Yucatan chachalacas. The Maya, not knowing the relevant science, might abstain on what the essence is or how you could tell when something had it. Another sensible answer would be: "No". If species are natural kinds at all, then it's likely that species boundaries are in fact determined by ancestry, and the Maya are here being given the relevant information. They might already think of ancestry as crucial. Or they might expect that whatever the essence consists in, ancestry would be a good guide on the subject. One would not predict, however, that they would answer "yes". Given that there is an important difference between chachalacas and twin chachalacas, the possibility that the twin birds are not chachalacas, but merely resemble them, ought to be salient. The Maya subjects, then, would not be expected to answer "yes".

The results were: "yes" 20, "no" 17 and "maybe" 4. These data certainly support the motley hypothesis and count against the natural-kind hypothesis. Of course, they are not conclusive. The issue turns on the best overall account of Maya cognition. Other things the Maya think would also need to be taken into consideration. But surely these data, which tap Maya intuitions about animals, should be given considerably more weight than Putnam's intuitions about Oscar's "water" concept and Kripke's intuitions about medieval "unicorn" concepts.

BIBLIOGRAPHY

Atran, S., 1999, "Itzaj Maya Folkbiological Taxonomy: Cognitive Universals and Cultural Particulars" in Medin, D., and S. Atran (eds) *Folkbiology*, Cambridge MA: MIT Press.

Carey, S., 1985, *Conceptual Change in Childhood*, Cambridge MA: MIT Press.

Evans G., 1973 "The Causal Theory of Names", *Proceedings of the Aristotelian Society*, Supp Vol 47 187-208.

- Evans, G., 1982 *The Varieties of Reference*, Oxford: Oxford University Press.
- Fodor, J. A., 1987 *Psychosemantics*, Cambridge MA: MIT Press.
- Fodor, J. A. 1991 "A Modal Argument for Narrow Content", *The Journal of Philosophy* 88: 5-25.
- Fodor, J. A. 1994 *The Elm and the Expert*, Cambridge MA: MIT Press.
- Kripke, S., 1972 *Naming and Necessity* in D. Davidson and G. Harman (eds) *Semantics for Natural Languages*, Dordrecht: Reidel 253-355.
- Kripke, S., 1973 unpublished, John Locke Lectures, Oxford.
- Millikan, R., 2000 *On Clear and Confused Ideas*. Cambridge: Cambridge University Press.
- Peacocke, C. A. B., 1992 "Externalist Explanations", *Proceedings of the Aristotelian Society*.
- Pinker, S., 1994 *The Language Instinct*, Penguin.
- Putnam, H., 1975 "The Meaning of 'Meaning'" in K. Gunderson ed. *Language, Mind and Knowledge*, *Minnesota Studies in the Philosophy of Science* 7. pps.
- Segal, G., 1989 "The Return of the Individual", *Mind*.98, 39-57.
- Segal, G., 2000 *A Slim Book about Narrow Content*, Cambridge MA: MIT Press.
- Segal, G., 1997, "Content and Computation: A Critical Notice of Jerry Fodor's *The Elm and the Expert*", *Mind and Language*: 491-501.
- Segal forthcoming a "Intentionality", in Jackson., F., and M. Smith (eds) *The Blackwell Companion to Contemporary Analytic Philosophy*, Oxford: Blackwell.
- Segal forthcoming b "Keep Making Sense", in Crane, T., ed., *Contemporary Philosophy in Focus: Jerry Fodor*, Cambridge: Cambridge University Press
- Segal, G., and Sober., E., 1991 "The Causal Efficacy of Content," *Philosophical Studies* 63, 1-30.
- Williamson, T., 1998, "The Broadness of the Mental: Some Logical Considerations" in Tomberlin, J., ed., *Language, Mind and Ontology, Philosophical Perspectives* 12, 389-410, Oxford and Boston: Blackwell.

White, T. H., 1954 *The Book of Beasts: a Translation from a Latin Bestiary of the Twelfth Century*.
London: Jonathan Cape.