failures of reference and failures of meaning
manque de référence et manque de signification

presidential address / allocution présidentielle ACPA May 2014

[lightly revised after ACPA presentation]

abstract: Philosophers once wielded criteria of intellectual respectability against over-ambitious or wacky theories. These were themselves overambitious, and look wacky now. I attempt, cautiously, to reconstruct something like a criterion of cognitive significance by combining causal accounts of reference with the division of intellectual labour.

précis: Il y a un temps les philosophes employaient des critères pour éliminer les speculations trop ambitieuses ou farfelues. Nous avons renoncés ces critères. J'essaie, très tentativement, de reconstruire un critère de signification cognitive, en combinant les théories causales de référence avec la division de travail intellectuelle.

DRAFT: May 014

epistemic-semantic teeth

In living memory, just, there was a fearsome beast called logical positivism, a kind of sly serpent that would slay its victims by surrounding them and then ignoring their cries, insisting that since they were neither true nor false they could make no intelligible appeal. After the world became bored or disillusioned with positivism, its place was taken by falsificationism, with different teeth and more syllables. Falsificationism was a kind of sphinx: after cornering victims it would force them to answer a riddle, "when would you be wrong?" If they had no answer it would toss them into the abyss. If they said "in circumstances C", it would smile quietly and

set about finding cases of C. Before and after both positivism and falsificationism there was a pack of little evidentialist carnivores, from the old days of Descartes and Locke to the new Bayesian era. Their teeth were not so big but they could inflict serious damage on a doctrine that had wandered too far from safety, leaving it vulnerable to doubt, refutation, and other causes of extinction.

These are all different, homing in on different vulnerabilities of their prey. Positivism is a semantic doctrine, though a semantics with epistemic roots. In generally Kantian style it aimed to find grounds for withholding meaning, intelligibility, from claims whose ambitions had taken them beyond possible evidence. Falsificationism is a suggestion about intellectual respectability. It excludes from science, and by implication any systematic inquiry, any view that could never be definitively refuted. (Why look for reasons to believe a claim, it suggests, if the deck has been stacked so that the inquiry could never come up with a negative answer.) But all three can be used for similar purposes. All can be used to give systematic reasons for rejecting large classes of [claims] on the grounds that they cannot be part of any legitimate inquiry.

Evidentialism is still with us, and honourably so in spite of the difficulties of formulating it [plausibly.] But neither positivism nor falsificationism is a live part of contemporary philosophy. Nor should they be: each has insuperable problems, that were once familiar to all philosophy students. (Some of them stem from an epistemic orthodoxy and a semantic one, epistemological holism and causal theories of reference, [both] of which will play a role in what follows.) But the range of views that it would be nice to be able to dismiss out of hand has not diminished. Apart from cranky religions and your valued colleague's philosophical views, there are still too-sweeping theories of history and society and the Place of Humanity in the Universe. Where have all the positivists gone now that we have homeopathy and climate change denial?

In this paper I recreate something like positivism, in terms of somewhat more contemporary non-epistemic ideas about meaning. It will follow from what I say

that some claims do not have a truth value that can be supported or [refuted] by evidence about their content. But the view I shall describe will also be [consistent] with a post-positivist orthodoxy, that is evident in the writings of the later Wittgenstein. That is that there can be no mechanical formula for separating sense from nonsense, no grand gesture with which one can sweep away whole piles of intellectual rubbish. The most that we are going to get is a technique, with which we can investigate intelligibility, so that the question of whether a claim has a truth value is, like the question of what its truth value may be, subject to evidence.

I begin with a background to this second, less confident, side of things.

the division of epistemic labour [FTN: SM: Witt]

Economists have long been comfortable with the idea that different people do different things, are generally not capable of doing what one another does, and that this specialisation is a source of the power of an economy. (In something like its modern form it is found in Mandeville in the 1700s. Plato had a sort of an anticipation of it, with qualifications.) There is an epistemic form, which has economic implications. Different people know different things. In particular they do not have the knowledge needed to fill others' economic roles. Moreover, they do not have an [effective] second order grasp of what knowledge is distributed in what ways. I, for example, have no grasp of many centuries-old technologies. How do you make a wooden chair, starting with a tree? What special expertises are needed and how long does it take? What wood is suitable and how long does it have to have dried? What glue, what tools? If I don't understand about chairs then my ignorance about cars and computers and l.e.d. lights is overwhelming.

The second order failure is of practical importance. If half of us died from a war or a plague, it would be up to the remaining half to reconstruct how it all worked, who did what and how they did it. Could they do it? Most likely not. It might take centuries, and then it would be a different distribution of roles and expertise. (Knowing how Japan did it in the late 19th century would be important information.

How similar was the array of expertises to that of the European and American models? The difficulty of making working capitalism in a fundamentally similar post-soviet eastern Europe is relevant, too.)

My focus now is on the general failures of higher order knowledge, KK in the general form of knowing who knows that p, knowing who, if anyone, knows how to A, and more subtly knowing which combinations of other people's knowledge is used to make E happen. I am taking it as clear, but under-appreciated, that any individual's second order knowledge of these kinds is very limited and fragmentary, but that the fragments often fit together to enable collective accomplishment. A jigsaw puzzle where the pieces are all on a plane and do not allow an oversight of how the further ones fit together. In the purely knowing-how form the failure is of knowing who has various congruent pieces of expertise to back up one's own opinions. For example, I know that penicillin is effective against a range of bacterial infections and like all antibiotics ineffective against viral infections. (As I write this I have to suppress the impulse to look up penicillin facts so as to appear less ignorant than I really am. But that's the point: and I'm a well-informed person.) But I do not know which bacterial infections it is (still) effective against, in what doses, how it is manufactured, what its mechanism of action is and whether it uses a similar mechanism against different bacteria, what yeast (?) it was originally cultured from, how it is transported in the body to whatever are the locations where it can interfere with an infection, and what background conditions are needed for it to function. And I do not know who knows these things, or even if there is anyone who does, in the sense of a single expert who has all the information. Is there a single person who can identify a set of people one of whom knows each of these? I don't even know the answer to that.

As a result, I cannot assess the force of inferences to the best explanation that appeal to the efficacy or limits of penicillin. (I'm not in a position to know when an explanation is better than alternatives.) I cannot gauge the force of evidence for and against claims about its efficacy against various diseases, as I would have to know whether the experiments control possible interfering factors suitably. So when

an expert -- my doctor, an experimentalist reported in the media, the author of a popular science article -- makes a claim about penicillin I usually have to take it on faith that the expert knows what she is talking about. To identify experts as such I do not give them exams (the idea is ludicrous) or judge the plausibility of what they say myself, but go by two things. I use (a) the endorsements of other experts whose knowledge I am in some position to grasp, exploiting an overlap between the people who are experts in one topic, which I can judge, and the people who have first and second order knowledge of topics that are further from my ken. {MORE?} I also use (b) a very limited third order knowledge, that some people know which others know who some of the experts are. For example I trust the authors of an encyclopedia article on antibiotics to cite works on penicillin whose authors would be able to consult with the right people to fill in some of the gaps I have mentioned. I feel that on scientific topics, at any rate, though my ignorance about who knows what is often abysmal, I can usually with enough effort figure out who the experts are. But it takes effort, too much for it to be worth doing regularly.

the division of semantic labour 1: names

There is a big theme here, worth developing at length. The special case of interest now is the distribution of semantic knowledge. My claim is that it is a typical case of the situation I have sketched, in that an individual's use of a word will only refer to something via the beliefs and actions of others, often many others, and the identities of these others and what knowledge enables them to serve this semantic role is often not available to the individual. As a result, I claim, it is both easy to make assertions that are semantically (referentially) deviant and also easy not to know that one is doing so. This limited claim should be already be clear: that if semantical facts are social they require knowledge of the intentions and beliefs of others that can be very hard to acquire, so that confidence either that a sentence makes sense or that it does not will often be very hard to establish.

First a toy case. Assume that the causal account of the reference of proper names is generally correct. Then when I use a name, for example ["Stephen”] then I am typically uttering it in a conversation in which others will take me to be making a

claim about some person whose doings and attributes are relevant. Which person this is, whether my son Stephen Morton, my grandfather Stefan Garibian, my colleague Stephen Savitt, the Prime Minister Stephen Harper or the physicist Stephen Hawking, will depend on the conversation. To contribute to the conversation and say intelligible things about the Stephen in question, I have to know who the other participants are and what they took the topic to be and who they had in mind. Often the immediate participants are themselves linking their use of the name to others' use of it in other conversations, so the chain goes back. All quite demanding of my mind-reading, though it happens automatically. This much is thoroughly familiar, though not universally accepted. Consider, though, some of the ways it can go wrong.

There can be changes of reference. A name can originally refer to one thing, and then get used so persistently to refer to another (perhaps saying false things about it) that it comes to refer to another. (Evans' *Madagascar* case; there are thousands. *Canada* for that matter.) A contemporary user of the name might become aware of the original use but not of the referential intentions of crucial intermediaries (though she would know that there was a chain, with intermediaries). Then she would use the word to refer to the "wrong" object.

Note that in cases like this there will be a period when neither the original nor the later reference of the word is the one it definitely has. Worse, the reference may be one thing for one user and another for another. Perhaps for some speakers there are some conversations in which it refers to one thing and others, at roughly the same time, in which it refers to another. (I think I can make cases.) Perhaps even there are conversations in which it refers to one thing for some participant and another for another.

There can be contrary pulls. Some factors may tend to one referent and some to another. These can be got from cases of change of reference, but for a different kind of case suppose that some of us are talking about time in physics and referring to the views of "Steve", that is professor Savitt, except one who has the pretension

of being close with Hawking. So when this guy says "But Steve told me .." he has him in mind. I, joining the conversation in mid stream say "But Steve thinks nothing like that". How is my assertion to be evaluated?

There can be too many referents. For a case that doesn't fall under the previous headings, suppose that we are using a rare and unique name, say "Oenonous". We have views about this person, and pass them to one another. But our referential chains go back to an original naming in which two babies were called "Oenonous", simultaneously, and the referential chains then fused, so that if we could trace the reference back we would find it splitting, and giving us two individuals with different attributes. (Think of Dionysius the areopagite.)

There can be too few referents, that is none. We can be using a name that was just made up, some time ago, as a joke.

There can be confusion about whether a word is a real reference-aiming name. A child who overhears her parents saying "Santa may bring her that telescope" is misunderstanding their intentions. They were not using "Santa" as a name of a person who may have attributes that are ascribed using the name, though they pretend to when talking to her. Some philosophers' interpretations of mathematical and theological language suppose that adults make subtler versions of the same mistake.

When things go wrong in any of these ways, there can be failure of reference. A person may use a word sincerely and naturally, employing the same amount of information about the ways others are using it, and their intentions in doing so, that people do in the innumerable cases in which things do not go wrong, but not succeed in referring to anything. In some instances of the last two of the sources of failure I listed, more properly semantic knowledge or sophistication might have prevented the problem. The same is true in classical nonsense cases, in which a word is used without the connections that give it meaning. But this makes failure of reference, too. When the cyclops says "No one has blinded me" he wants to refer to

Odysseus but he does not succeed. And there are no Jabberwockies. The failure of nonsense to do the work of sense comes from the more basic failure of reference.

In all these cases, too, a person will often not know whether her attempts to refer have succeeded. She will frequently refer (accomplish reference, to use the terminology I have developed elsewhere) without knowing that she has. And she will sometimes fail to refer unknowingly too. ('Frequently' and 'sometimes', just because normal speakers succeed more often than they fail.) One central reason for this is the semantic division of labour. For you to refer many others must play their parts, and though they usually do this correctly, the bad cases are distributed among the good cases in a way that is hard to ascertain. Combine this fact with the definiteness of reference failure and we see how a "positivist" ubiquity of semantic failure can coincide with a "Wittgensteinian" lack of a simple criterion for telling when failure has occurred.

things that can go wrong

It is essential to the analysis I am giving that many things can go wrong in connecting a particular use of a word with an individual referred to. Keep three in mind, which correspond to the three toy examples of defective thinking below.

(I) *too many referents*. See the "Stephen" example above.
One might want to describe some such cases in terms of multiple or divided denotation, and then assign a truth value with asupervaluational or other similar device. My instinct is to take multiple denotation of a proper name to be not denotation at all. (The name has turned into a predicate.) But it does not matter. What matters is that something has gone wrong. The speaker's semantical presuupposition is false and the reference is nothing like what she takes it to be.

(II) *no connection* See "too few referents" above. [Mickey and Santa examples. Would be better to use an example of someone taking a conversation about Simpsons" characters as being about real neighbours.]
Again one can argue for truth, but the important point is that the process has gone wrong: a semantical presupposition is false.

(III) *broken links* One use of a name is deferential on another and that on another, and so on. Some of the links may have no connection with any referent, as in (II). But, also, some of the *uses*, or the names providing the links, may not exist in the required form. A crude example: I can take my use of "Bart" to co-refer with yours, and you may take yours to co-refer with your sister's use of it: but she was just belching.

These cases might be assimilated to (II), but I think it helps when extending the picture to cases involving predicates to keep them apart.

the division of semantic labour 2: predicates

On some accounts of some species words and substance words ("Tiger", "water") they function in a way directly analogous to proper names. The conclusions I drew just above will apply directly to them. Instead of developing this analogy, however, I will explore a different line, which I think promises a deeper insight into the ways we can make less sense than we intend to. This is the [method] of *distributed Ramsey sentences*. [Mole note!]

Suppose that you are explaining a theory, for example the theory that nucleic acids from outer space originated life on earth, to someone who not only does not subscribe to it but does not have a good grasp of the concepts involved, nor of those in terms of which you are at first inclined to explain them. One strategy you could follow is to list all the connections between the relevant concepts and the evidence that would support the theory and which it would explain. (X causes Y; when Z is found W usually follows; whenever E occurs we can find some Ss.) Then you say "the idea is that there are kinds and properties related in these ways. See how if these exist then the data are explained. " This enormous tangled second-order formulation is the Ramsey sentence for the predicates in question. {Ramsey, Boehnert, Lewis} All going implausibly well (see below) the result will be that your friend sees the explanatory force of the theory, can compare it to rival explanations of the same data, and can compare the ways the theory understands the terms it uses (X, Y, A, W, E, S,...) to her own use of the same words. She may say things

like "this isn't what I take fitness to be, but in its own terms I see how it is supposed to work". At the end of the process you will have explained to yourself, too, what the content of the theory is, in a way that is relatively independent of your earlier uses of its terms, perhaps before you came to understand the theory. Indeed, you may only then realise that the theory understands its terms in a different way to rival theories and to other similar-sounding theories that you have considered.

There is a tension between accounts along these lines and accounts that treat terms in theories analogously to proper names. Suppose that the term "fitness" when first introduced is a synonym for "the number of descendants of an organism". But suppose that the theory of extra-terrestrial nucleic acid makes claims about fitness that are inconsistent with this reading. (Perhaps it makes fitness proportional to the number of strands of nucleic acid that are found in an organism's descendants. So multi-cellular descendants would count for more than unicellular ones.) Then on a simple-minded version of the causal semantics the ETNA theory would be false, although there may be a candidate for the denotation of "fitness" that, together with suitable denotations of other terms, satisfies the requirements of the theory. [I will not try to resolve this tension, but instead will argue that on this semantics, too, undetected failures of reference ought to be common.]

People are rarely going to formulate theories to themselves or to others in terms of such Ramsey sentences. There are many obstacles. There is the large number of predicates that a full R-sentence will need, connecting the predicate in question to all the others with which it is theoretically or observationally connected. Then there is the greater mental burden of thinking with and communicating with second order quantifiers rather than free-standing predicates.

These are consequences of a greater problem, and pale in comparison with it. The R-sentence has to be a single sentence, with second-order quantifiers binding all the occurrences of the relevant predicates, and whenever there is a connection

between predicates it must show up in the sentence. These sentences will be enormous, and it is an open question about human cognition how many sentences there will normally be. (How compartmentalised our thinking is, how interwoven the web of belief.) Perhaps there is only one: all a person's concepts and beliefs rolled into a single great intricate network preceded by "this is satisfied". In any case, if the Ramsification is carried out completely the results will usually be completely unmanageable as instruments of communication and thought.

But something Ramseyish is what is needed. After all, we frequently use a word in different senses in different theoretical contexts. "Energy" in "the Hamiltonian gives the energy of a system in terms of a vector field" and in "if you start the marathon at that speed you'll run out of energy before half way" are hardly the same word at all. So we want to be able to separate them by saying something like "there is a physical quantity that when incorporated into Hamilton's equations gives the trajectory of a system", and "there are resources that people use when they move their bodies, and if they don't conserve them they get depleted". This keeps the two apart, although since there are connections between the two concepts a really full exposition of the two should link them, probably in the form of a complex R-sentence incorporating both energy-in-physics and energy-in-human-activity as different second order variables, with additional clauses connecting the two.

The result is an array of limited, manageable, informally stated Ramsey sentences that we use to explain what we mean by a term, and what connections we take as essential and less essential to it. In terms of this we explore which terms refer and co-refer. It is crucial here that the array is usually held communally. Since most people are not capable of producing even the limited R-sentence for the use of most terms in a particular theoretical context -- how many can state or even recognize Hamilton's equations? -- they rely on expert others who can add essential details to their scraps of theory and connect what they know about a particular term to others, including others that they have no grasp of. 99% of those who cannot state Hamilton's equations have no idea what the distinction between kinetic and potential energy is. Moreover there is usually a number of experts, each with full

grasp of part of the content. A physics teacher may well be at sea on the continuity conditions presupposed by Hamilton's equations, but the mathematician who knows this may know nothing about the energy functions of typical systems. A textbook writer may be able to combine both, but not by being fully up to speed on either. Instead, she will allude to the restrictions that are needed or the ways the theory can be applied, giving some details and deferring to the specialists. Put physicist, mathematician, and textbook author together and you have something like full expertise, perhaps even the materials for a real R-sentence. Put the non-expert user of "energy" in touch with the three experts and you have the R-sentence that she needs. But it is not available from what she thinks or says.

(I believe these points combine well with a Burgean externalism about meaning. But that is for another occasion.)
[I also think that these points point to a resolution of the tension between the R-sentence approach and the causal approach. People using a term intend (believe, deny) the simple unquantified assertiojn using that term as a predicate (not as a second order variable.) When they owe their grasp of the term to others – which means usually – the R-sentence that expresses the content owed by speaker to expert determines its reference. It is like the first order quantifications "there is someone who speaker C has in mind in this assertion" and "there is a baby who we will know as n" which may determine then reference of proper name *n* for speaker A. Note that the expert also believes and communicates the unvarnished unquantified sentence. If this is right the distributed quality of theoretical belief is not simply a concession to human limitations but an intrinsic aspect of their referential power.]

Many things can again go wrong. The innocent user of a term may misidentify the experts. She may treat two uses of a term as calling for a common R-sentence when in fact they have only remote and inessential connections. She may misidentify the R-sentence that is appropriate for her use of the term. She may find that as R-sentences approach the complexity that spells out the content of their concepts they come to be too complex for her to grasp. [See the toy cases below.]

And all of this is before we come to questions of satisfiability. There may be no properties (quantities, kinds) satisfying the sentence, and even a savvy user of a term may not know this. One fundamental reason might be hidden inconsistency. The R-sentence may entail assertions that, if she were to think about it long enough, she would see to be irreconcilable with her conception. Of most interest here are the problems that in principle could be discerned by a super-intelligent agent without empirical investigation: misidentifications and complexity. These define situations in which a person's grasp of a term fails to link up to a specification of what the world must be like to provide a referent for the term. They are like the problems that await proper names on a causal account of these, but they are more varied and apply to a greater variety of terms. And -- essential to my themes -- there is no simple procedure for preventing them. A clear-headed, honest, scrupulous thinker can easily think in terms which a deeper analysis, which in general it is not in her interest to carry out, will show that her words are empty.

Of course it often happens that a clear-headed, honest, scrupulous thinker has a thought that is simply false, and in some such cases, according to some philosophers, the reason is that a term fails to refer. Different philosophers have different opinions about when failure of reference leads to falsity and when to lack of a truth value. I'd like to avoid getting into these disputes if I can. Instead, I shall operate with the following distinction.

> (a) cases where although a term does not in fact refer, there is possible evidence which would suggest that it does have a reference. (And which does not essentially change the kind of term it is: I'll leave this vague.)
>
> (b) cases where there can be no such evidence. (So it is built into the kind of term it is that it does not refer.)

Note that this introduces an epistemic consideration into our semantic [criteria.] Note also that a person might misunderstand which of these a case is. (Trivial example below.) The distinction between (a) and (b) can be illustrated with two contrasting demonstratives. If I point to my neighbour and say "this chair is

mahogany" then the demonstrative lacks reference - there is no chair there - and there is not going to be any evidence suggesting that that person is a chair. (There can be evidence that there is a chair where I am pointing, but that is different: I'm pointing at the person.) So that is a (b) case. My reaction to such cases is that the assertion lacks a truth value, but I won't insist on it. On the other hand if I point to a screen and say "that cat, behind the screen, is Siamese", when there is no cat there (perhaps there was a moment before) then evidence could turn up that there was a cat there after all. That's an (a) case. My reaction is that the assertion is false, but again I'll try to make nothing depend on this: it is (a) versus (b) rather than truth value or not, that matters. [King]

With proper names the classic example of (a) is "Vulcan". We could still learn that there is indeed a planet having the effects for which Le Verrier supposed the planet he named "Vulcan" was responsible. (It would be very surprising, to be sure.) The classic example of (b) is "Santa Claus" or "Mickey Mouse". These names are introduced as fictional, and it is to misunderstand their semantic status to think that we could learn that they name something real. (If we learn that someone lives at the north pole and distributes presents to obedient children on Christmas eve, then we have learned not that Santa exists but that someone else fits the story. Compare Kripke on unicorns.) [Mousavian] A child can of course not know that this is a (b) case, because she has been deliberately misled about the semantical status of the name by adults.

So the focus in this essay is on (b) cases. And especially on cases where the person using the word is wrong about whether it refers to anything, and thus, it being a (b) case, whether there can be relevant evidence about the truth value of assertions involving it. The division of semantic labour gives models of how this can happen. I have just given a trivial one: a child is led to believe that "Santa" refers to a real person. This is a division of labour case because it would not happen if the child's use of names were restricted to names of individuals she had encountered herself. Instead, she relies on the honesty of adults, which in most cases is [well-founded.] For more interesting, but still simplified, examples see the next section.

three toy cases

I now briefly present three toy cases to illustrate the kinds of failure my account imagines. They are toy and they are brief: very simplified models of ways in which people though taking themselves to refer may fall short. They have to be very simplified versions of real cases, given my emphasis on the difficulty of knowing when and how things may have gone wrong. One of the several ways in which they are simplified is their exclusive focus on the semantic/referential dimension. In real cases this is hard to separate from the epistemic dimension, so that two people who agree that a position is absurdly formulated may disagree about whether it is short of truth-evaluability or whether on the one hand it is hard to see that there is or could be evidence for it. (More subversive of belief than atheism is the suggestion that both the assertion and the denial that gods exist fail to get to the point of intelligible truth, falsity, or "real" belief, disbelief.) I shall ignore this important issue (but see the remarks in the final section of this paper.) [Is the second half of this para now badly placed given remarks elsewhere?]

(I) "*energy transmission*" Adherents of a small healing discipline believe that they can transmit their own health to others by summoning the energy that is manifested in their well-being, sometimes by readying themselves as if for physical effort, and thinking of the other person. When asked what energy is they say it is the same as the biological energy that allows physical movement and the kinetic energy that moving bodies have, and when asked how it is transmitted they say it works the same way that a determination to move a tired limb goes from mind to limb. They appeal both to the folk psychological ideas that you do something mentally which makes you do something physically, and that you can make someone think something by thinking and acting yourself (for example by smiling at them), and the physical idea that there is a quantity that is conserved and transmitted when one kinetic event causes another. The energy of will turns into the energy of your own or another's body." "You are active in your mind" they say "and you send that energy to the suffering person."

Diagnosis: The view *might* have been semantically evaluable, if we could separate

this concept of energy from the one found in physics. Then it would have a chance of being intelligible but probably false, were it not for the dual reliance on energy in dynamical systems and motivational energy, the unholy combination of folk psychology and physics. So the situation is more dire than just that "energy" as they use it, and "transmit" - "send", "give", "beam" - refer to nothing. But given this simultaneous reliance on two incompatible uses of the terms, we have no coherent way to fix what they are discussing, even to the extent of saying something false about it. It's a (b) case.

(II) *"school mathematics"* (II) a TED-talk guru offers an explanation why girls do wose than boys at school mathematics. Postulates a property of the corpus callosum in boys' brains that makes them better at quantitative/symbolic reasoning. Not explicit about what this might be, uses scattered conjectures about possible to explain the "fact" that girls perform less well.

diagnosis: The fact is a myth; recent data suggests that girls have always done as well as boys in school math. (Sexist confession: shocked to find this surprised me.) So the only reference-fixer the guru gives, that the mystery property is what explains the gender difference, fails to connect with anything. It's a mahogany chair case.

(IIA) "*race*" [left out of CPA talk: doesn't add much to I] A group of people share the attitude that they are members of the same subset of humanity, and take the rest of the species to be divided into a finite number of discrete non-overlapping other subsets, which they call "races". They associate each such race with a set of character traits and capacities, though they defer to one another for the list of these, and consider them to be hereditary, leaving it undetermined how they are passed between generations and assuming that "scientists" know how it works. They all take the races to be recognizable by sight. They believe the same of other species, so they take poodles, Rottweilers, and spaniels to be races of dogs. They assume that evolutionary biology and Mendelian genetics support their views, and they say "there are races of humans, just as there are breeds of cattle and dogs,

and subspecies of ants", appealing to the biological notion of a subspecies.

Diagnosis: There is no biological property with all these characteristics. To a first approximation the position is simply false. It says "there are properties such that ..." where the ... cannot be filled in to get anything true. But to say just this ignores the deference to biology and the assumption that science can fill out the details to make a coherent theory. In fact this cannot be done, and the inheritance of physical characteristics in a subspecies has nothing to do with any feature of visually identifiable groups of people, and the biological analogy is based in part on special features of subspecies bred by humans for their purposes. So our racists think they are borrowing their reference from biology: the idea is that biologists can put together a true Ramsey sentence that unifies theeir various first order views. But there is nothing in biology that serves the range of explanatory functions that would be required. The aim of trcing reference back to a scientific discipline has to fail.
[modification? Diagnosis: To a first approximation the position is simply false, since there is no attribute to fit its term "race". But this ignores the promissory note that when details are filled confirming facts emerge. In fact neither genetics nor animal husbandry will give the details that are needed, so the position survives by oscillating between them.]

(III) "*enlightenment*" A group of meditators take their activities to be steps towards enlightenment. They characterise this loosely, especially to outsiders, as escape from the cycle of rebirth, but they also say that this is just the kind of too-verbal and too-conceptual formulation that one abandons on the route to enlightenment, and is hard to reconcile with the insight that the self is an illusion. Instead, in serious discussion they rely on (i) the sequence of enlightened leaders --- each assured by the previous one that what he has attained is indeed enlightenment, and that it is what he has been seeking -- tracing back ultimately to the Buddha. And (ii) the assumption that the state is self-recognizing: when you have it you know it. But in fact they are often in doubt about their status, and so rely on the diagnoses of their elders, particularly to assure them that they are not there yet and must stick to the path. So (i) is the central factor.

Diagnosis: In fact, the sequence of enlightened sages is broken, and at some point contains names of people who never existed. The chain of assurances peters out, first into unverifiability and then into myth. As a result, the only effective characterisation of enlightenment our meditators have, in fact the only way their term might refer to anything, is in terms of a chain that comes to nothing. (They think they have a *Mars* case but they don't even have a *Vulcan*, since it's a *Santa*.)

I should add that I am not saying that people like those in the "energy transmission" case do not have healing powers, or that no one reaches a special state by meditation. Or even that there may not be importantly different subpopulations of humanity (though it is incredible that they should be distinguishable by sight.) It is just that what these imaginary people say to explain their views is semantically defective. Crucial terms do not refer, and do not refer for reasons that go deeper than the failure of nature to provide suitable referents. They do not refer because the links that would connect them with referents are tangled, incoherent, inadequate. (That is (a) cases versus (b) cases again. There are obviously many (a) cases too. I'm inclined to take homeopathy, for example, as meaningful but false. It postulates molecular memory but there is no such thing.)

Also, and this is crucial, I am not saying that anything like these diagnoses, bringing in accounts from outside the people's thinking about how their referential processes are functioning and what there is to refer to, could be used by the people themselves to see that their opinions are hollow. No, from the inside possession of such an opinion feels no different from possession of any other. The links to reality are faulty, but ignorance of that is of a piece with any of the other innumerable things of which we are ignorant. So the semantic failure in 'my' cases is like a paradoxical consequence of externalist accounts of singular thought: something in the person's thinking functions in the way that a real content-bearing thought does, but there is not in fact a content. [Jeshion] (And you can use this resemblance as a quick handle on the central idea of this paper: grandiose failed theories have a lot in common with failed singular thoughts, except the failure is typically spread over a network of thinkers.)

the central issue

I have a suggestion about the central issue here. I think there is a technically hard problem which is also conceptually confusing, that has a lot to say about these issues. Begin with an example, that at first sight has little to do with semantic deviance.

Example:
A diamond has been stolen and the only suspects are unmistakeable Mike and doubled Dinah. they're so called because Mike has gelled hair with purple spikes plus a green beard and tatoos of the text of the tractatus across his forehead, while Dinah is a very ordinary looking person who happens to have an identical twin. the jewellery store clerk was putting valuables in the safe when the robber pointed a gun at him. he scribbled down the name of the robber and slammed the door shut, before collapsing.

The clerk is still in intensive care and we don't have the combination to the safe. but we do know that there is evidence in the safe -- the paper he wrote the name on -- about who the criminal is. if the paper says "Mike" then there is a 0.999 chance that he is the criminal. (I take that figure to give the force of the evidence.) if the paper says "Dinah" then the figure is 0.3. It is crazy to want somehow to average or weigh the 0.999 and the 0.3. Instead we have to say just that it is either one or the other but we don't know which.

(I expect there are variations where there are figures for weighing between the two, but still when applied the figure doesn't give anything like how strongly whatever is on the paper supports either hypothesis. I guess it gives a figure for how confident you should be that the evidence supports H. Or computed differently how confident you are now that eventually you will be confident -- to some degree? -- that H is supported.)

The underlying issue is this: often one has evidence about the existence of

evidence, often about whether it tends to one hypothesis with one strength or another with another. Factoring this meta-evidence into the first order evidence in order to see which hypothesis is better supported, and how much, is hard. I'm not convinced it can be done in a uniform way. (It's the topic of "meta-analysis" in statistical inference, that I am very suspicious of while knowing very little about.)

The connection with the present issues is in a special case. Sometimes the meta-evidence concerns whether there is or can be any evidence at all. Semantic facts are natural facts, and we rely on evidence about them, which can be hard to obtain and assess. So general conclusions about knowledge can be transferred to the particular case of semantic knowledge. Very often we have some reasons to believe that someone's use of a word has one reference, some reasons to believe it has another, and some reasons to believe that it has none at all. These can all compete in our interpretation of the person's words. So in order to come definitely to the conclusion that there is no reference we would have to resolve these tensions. And that is tough, which is why in examples like those I have been giving one is torn between evidential deficiency and semantic deficiency. (And why in order to make it uncontroversially semantic one has to make it artificially simple, thus losing the connection with the cases of intellectual disreputability that we want ultimately to address.)

What to do? I think we need tidy ways of presenting the first order and meta-evidence as such, with all their conflicts and their multiple destinations. Then we could say "here's grounds for thinking it is meaningful but false, here's grounds for thinking that it is a honest semantic mistake (and here's a reason why it might even be true, and one why it might be a culpable misuse of language." And leave it at that. (The "tidy" is the hard part.)

end: meaning and evidence

The claim of this paper is that we can reconstruct much of the force of severe old strictures on the dangers of careless theorising in terms of reference failure, using presently dominant accounts of reference. You can see this as restoring some old

positivist righteousness, but it also brings a characteristically modern lostness. For this account predicts that we will often be wrong about these matters. We will often take terms to refer when they do not, and it is hard to get and bring together the information needed to establish that a term does indeed refer. So while we might advertise part of the message as saying that a lot of what we say is really semantic failure, we ought to qualify this by noting that failure is not always told from success by the way it seems even to a careful and responsible speaker. They can feel the same. (In fact, though I have not given examples, there should be cases that look like failure but are in fact referentially successful. [Heavyside's step function?])

[NB how the uncertainty claim makes it doubtful that syntax is much help. Don't say 'grammar' [is this an example?!]]

Though they are similar in spirit, there are two contrasting strategies for undercutting overambitious thinking. The strategy I have been reconstructing focusses on semantic problems. The other strategy focuses on evidential problems. It would be valuable to understand better how the two strategies relate. I shall end this paper with three (not terribly profound) remarks about their connection.

First, there is the way that information about reference can shortcut evidence. Negatively, if one knows that a term does not refer then one knows that there is not going to be evidence for claims stated using it, at any rate in the (b) cases above. You don't have to do experiments to find out whether snarks are carnivorous. Positively, if one can bypass the chain of referential links and get information directly about an object then one has better evidence of its properties. A photo of Socrates would be worth such a lot more than "this book referred to a snub-nose mentor of Plato, and Plato said ... ". Do the negative and the positive hint fit into a comprehensive account? I don't know.

Second, though the focus has been on meaning rather than evidence, I have emphasised that semantic facts are natural facts, and we rely on evidence about them, which can be hard to obtain and assess. So general conclusions about

knowledge can be transferred to the particular case of semantic knowledge. Innocent semantic failure is a special case of ignorance. More detail about this might help with the comprehensive account just mentioned, but that is just wishful thinking at this stage.

Third, though in any real case we are unlikely to get a clean separation of epistemic from semantical failure, we can make a taxonomy of the ways reference can go wrong. And then in real cases we can say: these are the particular dangers of this way of thinking and presenting, so it would be reassuring to have an assurance that these particular dangers do not affect the claims at issue now. The account in this paper is just a tiny beginning of what would be needed.

Adam Morton

UBC

thanks to: Jonathan Bennett, Christopher Mole, Seyed Mousavian, Alex Rueger, and the audiences at UBC and Brock.