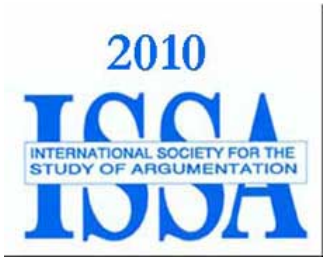
ISSA Proceedings 2010 – Agent-Relative Fallacies



1. Introductory

My topic is an issue in the individuation and epistemology of fallacious inferences **[i]**. My thesis is that there are instances of reasoning that are fallacious not in themselves, that are not intrinsically fallacious, but are fallacious only relative to particular reasoning agents. This seems like a peculiar notion. It would seem that if it was fallacious for you to reason a certain way, and I do the same thing, I would be committing a fallacy as well. Bad reasoning is bad reasoning, no matter who is doing it. But it is useful to ask: What would it take for it to be possible for there to be such a thing as an agent-relative fallacy? Here are two sets of conditions, the obtaining of either of which would be sufficient for the existence of agent-relative, or extrinsic, fallacies. Type One is that there are two agents who are intrinsically alike, molecule-for-molecule doppelgangers, one of whom is reasoning fallaciously while the other is not, due to differences in their respective environments. The other scenario, Type Two, is that there are two agents (who are not doppelgangers) who engage in intrinsically identical instances of reasoning, one of whom reasons fallaciously while the other does not, due to differences located elsewhere in their minds that affect the epistemic status of their respective inferences. I will attempt to demonstrate that it is at least possible for agents to meet either set of conditions, and that in fact some people do meet the Type Two conditions, so agent-relative fallacies are not only possible, but actual.

2. Type One Agent-Relative Fallacies

So could there be agent-relative fallacies of the first sort, Type One, in which one of two intrinsically identical doppelgangers reasons fallaciously and one does not? For such a thing to be possible, I think it is necessary that a strong thesis of internalism, or individualism, about mental content be false. Mental content internalism is the view that the mental supervenes on the physical, meaning that there cannot be a mental difference between two agents without a physical difference between them. Content internalism is a somewhat beleaguered position nowadays, in part because of Hilary Putnam's famous Twin Earth thought

experiment (Putnam 1975, *passim*) and arguments from Tyler Burge (Burge 1979, *passim*), in favor of content externalism. Putnam imagined a Twin Earth that is identical to Earth in every way, including Twin Earth counterparts of you and me and this podium, except that where we have water, Twin Earth has a liquid they call "water" that behaves just as water does, but which is not H2O – its chemical composition is XYZ. While the thoughts of a thirsty earthling turn to water, the denizen of Twin Earth has no thoughts about water, as she has never had any contact with water (i.e. H2O). Instead, her thoughts run to the stuff that is XYZ, which we might call twin-water. The earthling and her counterpart are doppelgangers (putting aside of course that we are composed in part of water) who behave the same way and make the same sounds, but they are mentally different, since one has water beliefs and desires, and the other does not (provided that content externalism is correct).

Suppose externalism is correct, and molecule-for-molecule doppelgangers can differ mentally. What one is thinking would not be an intrinsic feature of a thinker. Could this engender as well situations in which one doppelganger reasons fallaciously and one does not? Here's how it seems that it might. A widely noted feature of content externalism, for better or worse, is that it seems to undermine to an extent one's introspective knowledge of one's own thought contents. In particular, it seems to allow for errors about comparative content – that is, errors as to whether two thought tokens have the same or different contents, because the sameness or difference in content of two thought tokens depends in part on the respective connections to the environment those thought tokens have, and that's something that is unavailable to introspection, and about which it is easy to be mistaken. For example, one might suppose that one's assent to the sentence "water freezes at 0 Celsius" and one's assent to the sentence "water is now running from the garden hose" mean that one has two beliefs involving the same natural kind concept, *water*. But suppose that one has moved, unawares, from a water-environment to a twin-water-environment, and that the general belief about the conditions at which water freezes was prompted by experiences long ago with water, and is sustained by memories of water, while the current belief about what is coming out of the garden hose is caused by one's perception of twin-water. It is plausible in this circumstance to suppose, if content externalism is right, that one believes that *water freezes at 0 Celsius* and that *twin-water is coming out of the garden hose*, despite the fact that one takes oneself to be employing the same natural kind concept in both instances. Suppose

then it occurs to one to infer from those beliefs that *something is both coming out of the garden hose and freezes at 0 Celsius*. This will appear to be valid to the agent, an instance of lambda-abstraction (x is F; x is G; thus something is both F and G), but it will not be valid because the agent equivocates, using a term with different contents in the different premises, and trades on the supposed identity of content in inferring the conclusion. (I am taking it that the different meanings of 'water' in this argument are sufficient for it being equivocation even though in a fairly straightforward sense the subject seems to be guilty of no shortcoming with respect to her *logical* skills.)

The example might seem too fanciful, as it involves someone being switched unawares from Earth to Twin-Earth (and the notion of Twin-Earth itself is a bit dubious, as it may be physically impossible for there to be a substance that superficially is just like water but has a different molecular substructure). But Tyler Burge's version of externalism holds that an individual's thought contents can be dependent on the practices of the linguistic community to which she defers, and switching unawares from one linguistic community to another is not so far-fetched. For instance, the word 'billion' picks out different numbers in different English speaking linguistic communities. The US has always used the "short scale", on which 'billion' picks out 1,000,000,000 (ten to the ninth power, or a thousand millions). Although this short scale is becoming the dominant scale, there is a long scale according to which 'billion' refers to ten to the twelfth power (or a million millions – a trillion on the short scale). The long scale was operative in Australia, among other places, and is still used on some official documents. Suppose Suzy was raised partly in Australia (when the long scale was popular there) and partly in the US and belongs to both linguistic communities equally. Suppose further that Suzy doesn't know exactly how many a billion is, just as I – I must admit – do not know exactly how many is a googolplex. Just as I can have beliefs that employ the concept of *googolplex*, such as my belief that a googolplex is larger than a trillion, even though I do not know how many a googolplex is, Suzy can have beliefs that employ the concept (or *a* concept) of *billion* without knowing how many a billion is.

Suppose Suzy is living in Australia for the summer and reads in an Australian newspaper that "The US national debt is $13 billion" and she confirms this in her economics class at an Australian university. She comes to believe the (true) proposition expressed by that sentence. That winter she spends in the US and

there she reads about Bill Gates and his net worth of $53 billion, and she comes to believe that true proposition too. She defers to the experts and the rules in each of her linguistic communities, intending to mean by 'billion' whatever that terms means in her community. Now it occurs to her to put together her true beliefs about the US national debt and about Gates' net worth, and she concludes that Gates has more than enough to pay off the US debt (although of course this is not the case). As with the water/twin-water inference, one probably would be reluctant to question Suzy's logical acumen, but it looks like she equivocates (provided that Burgean social externalism is right), and she is open to at least some degree of reproach, for not making sure that she was not doing this. (Though I think you could construct examples where this linguistic shift is so subtle that she's not subject to any reproach at all.) And had both of her linguistic communities used the short scale, she could have had the same experiences and have been the same from the cranium in, but she would not have equivocated, as the premise expressed by "The US national debt is $13 billion" simply would have been false. So *what* she is thinking – and whether she is thinking fallaciously – is not an intrinsic feature of hers. ('Chicory' and 'football' are also examples of terms that have different extensions in different English-speaking communities, but which are similar enough that there is a potential for this sort of confusion.)

There are several ways of resisting this conclusion but I do not think any of them work. For instance, one might insist that because Suzy's inferential behavior indicated that she took the concept expressed by "billion" to be the same in each inference, it must have been the same concept each time. So there must have been a false premise, but no equivocation and no logical error. This has some appeal, as we are reluctant to judge this victim of the vicissitudes of travel as logically deficient. But this, it seems, is to reject content externalism in favor of some sort of internalist theory of the individuation of mental contents, an inferential role theory of some sort. So the first kind of extrinsic or agent-relative fallacy is possible on the condition that content externalism – a leading theory of mental content – is the case. And the sort of content externalism that must be true here is not necessarily as strong as the sort claimed by Putnam and Burge. All you need, I think, is that at least indexical or demonstrative thoughts – involving 'here', 'I', 'now', 'that' and so forth – are individuated in an externalistic manner. For example, from 'You said hello' and 'You smiled', it follows that you both smiled and said hello, only if 'you' picks out the same person each time (and perhaps that you have good grounds for supposing that it does as well). (I'm

assuming here that as long as the term is indexed the same way in each premise, or the same thing is demonstrated, and the agent is entitled to suppose that it does, then the conclusion follows validly. David Kaplan has argued against this, actually (Kaplan 1989, pp. 587-590), saying that the *potential* for distinct referents, when there are distinct demonstrations, creates the *actuality* of equivocation. So it is fallacious, on his view, even if the same object is demonstrated each time. This implies that one cannot deductively reason with premises using demonstratives, or at least not in a way that depends on the identity of the distinctly demonstrated demonstrata. I do not think this is a good idea, though, as the 'water' and 'billion' examples, and cases of two people with the same name, show that there is the potential for distinct referents in a much wider set of situations. I think this too narrowly circumscribes the sort of terms with which we can deduce.)

The possibility and actuality of Type One agent-relative fallacies thus depends only one a fairly plausible metaphysical claim about the individuation of mental thought contents.

3. Type Two Agent-Relative Fallacies

The second type of agent-relative fallacy is that an inference is fallacious for one agent but not for another, because of differences elsewhere in their minds that affect the epistemic status of their respective beliefs. This is to be in a way holistic about fallacies, maintaining that whether an inference is fallacious depends not just on that inference considered in isolation, but on the rest of the agent's web of beliefs as well. One way to illustrate this (and this example is due to my colleague Michael Veber) is to consider the case of *ad verecundiam*, or irrelevant appeal to authority. *Ad verecundiam* is committed when someone argues for a proposition by pointing out that some authority or expert has asserted that proposition, when in fact the proposition is outside the authority's area of expertise. Of course, it can be hard to say whether something falls within one's area of expertise or not, as expertise can be a matter of degree. Suppose I say that we should accept the claim that there is probably intelligent life elsewhere in the universe because scientists Carl Sagan and Stephen Hawking have said so. It would be a commission of *ad verecundiam* to accept a proposition that falls within the purview of science, broadly, just because some famous scientists have asserted the proposition, but it would not be if one had evidence that the proposition was within those scientists' area of expertise. So I take it that

whether the appeal to authority is fallacious or not depends not just on whether the cited experts are genuine experts on the matter at hand, but also on whether one has good grounds for taking them to have such expertise. Were I to defend a claim about string theory on the grounds that it was asserted by a stranger on the train, I would be guilty of *ad verecundiam* even if it so happened that this stranger were, unbeknownst to me, the world's leading expert on string theory. So it seems plausible that two people could make the same appeal to the same authority in defense of the same claim and that one does so fallaciously and one does not, because one lacks the right sort of evidence about the authority's expertise and the other has it.

I suppose that you could resist the claim that these two people with different evidence available to them nevertheless made the *same* appeal to authority, as adducing the evidence of expertise is *part of* the appeal to authority. If the appeal to authority really were the same for each person, then the one agent's superior evidence isn't playing the role that it would need to, in order to stave off *ad verecundiam.*

So consider another type of case. Various philosophers have theorized that particular forms of inference are fallacious – or at least that they don't confer justification on their conclusions. David Hume (arguably) thought this about induction, William Lycan and Vann McGee have argued that *modus ponens* (or at least some instances of it) are invalid, and Baas van Fraassen has argued against abduction (or inference to the best explanation). Let's take van Fraassen. He's argued that inference to the best explanation, or abduction, doesn't confer justification on its conclusions because – and this is just one reason among several – for any good explanation E of a set of data, there is an infinite number of equally good explanations of the data that are inconsistent with E (van Fraassen 1989, p. 146). Van Fraassen is a brilliant philosopher and he has evidence against abduction, but we shall suppose that he is wrong, and that inference to the best explanation is a legitimate way of inferring justified conclusions. Suppose further that while he tries to abstain from inference to the best explanation in his daily life, he frequently engages in it anyway. (C.S. Peirce, who introduced abduction to modern logic, thought that abduction was the first stage of all reasoning, and that nobody could avoid it.) Van Fraassen, for instance, receives a paper from a student that is a word-for-word duplicate of a paper published years ago by a notable philosopher, and infers that the paper is likely plagiarized, rather than

that the exact similarity between the papers is a matter of coincidence. So abduction is (generally) not a fallacious form of reasoning, van Fraassen engages in abductive reasoning on a daily basis, but he has a theory that abduction is fallacious and must be eschewed. What are we to say about the status of van Fraassen's own abductive inferences?

Well, they are not fallacious in the sense that they have a form that is particularly likely to lead to error. Presumably, van Fraassen is no more likely to fall into error using abduction than anyone else is; we will stipulate that. There is a question, though, as to whether he'd be epistemically justified in the conclusions he reaches through abduction, given that he has reasons to think abduction is no good. So for this sort of agent-relative fallacy to be possible – where an otherwise perfectly good inference is fallacious because the agent has evidence that it is fallacious but employs it anyhow – two things need to be the case. *One* is that it is sufficient for a truth-preserving inference to be fallacious that it fails to preserve epistemic justification. *Two*, it must be the case that if an agent has evidence that a particular sort of inference is fallacious but draws that inference anyhow, then she is typically epistemically unjustified in the conclusion that she draws. This would mean that the evidence that van Fraassen has against abduction would be a defeater for the particular abductive inferences he makes. If these conditions are met, then the van Fraassen abductive inferences (and similar cases) would be fallacious (even though they are just like yours and yours are not fallacious).

So, the first one: for an inference to be fallacious, is it sufficient that it be unable to deliver epistemic justification of the conclusion, even if the inference is truth-preserving? Well, the question of how to define 'fallacy' has proven quite difficult, and is necessarily beyond the scope of this short paper, so I will just point out that it is difficult to distinguish between fallacies and non-fallacies without bringing epistemic justification into it. Consider 'this entire throne is made of gold, thus the seat of this throne is made of gold'. This does not seem fallacious though superficially it is fallacy of division, and I think this has something to do with the fact that belief in the conclusion is epistemically justified by the premise.

The second condition: if one has evidence a particular inference type is fallacious, but one goes ahead and employs it anyhow, would one's resulting conclusions be unjustified? Let me point out that to answer 'yes' here is not to commit *tu quoque* (as when one says 'your argument in favor of vegetarianism fails, because you're

eating a hot dog right now!'); rather, a 'yes' answer would mean that evidence about one's evidence can undermine one's justification for first order propositions, as one must respect the evidence one has about one's evidence. So the situation is not just that one's beliefs are at odds with one's inference, but that one has evidence against the reliability of the inference that one is not properly respecting. To assert that if one has evidence that an inference type is fallacious, but one draws inferences of that form anyhow, then the inference is epistemically unjustified is perhaps to endorse the following *epistemic descent* principle (a principle moving from second-order epistemic claims to first-order ones**[ii])**:
(EDJ) If S believes with justification that y is unjustified (where y is an inference rule), and S believes that p only as a result of employing y, then S's belief that p is unjustified.

This is not to say that in order for a first-order belief to be justified, one must have any particular second-order belief about the first-order belief – surely children may have justified first-order beliefs even if they lack any second-order beliefs – but that one must *not* have a justified second-order belief that the first-order belief is unjustified. In fact, a stronger principle seems defensible:
(EDU) If S believes without justification that y is unjustified, and S believes that p only as a result of employing y, then S's belief that p is unjustified.

The idea here is that as long as one does believe that a particular first-order belief is unjustified, it would be unjustified for that agent. This is one strand of a broader view: defeaters themselves don't need to be justified in order to defeat justification. (For instance, although one is normally warranted in relying on her memory in forming beliefs about the past, one is not warranted in doing so if one is convinced that her memory is unreliable. This is so even if her reasons for thinking her memory to be unreliable are poor ones – that she believes it is sufficient to make her unjustified in forming beliefs about the past based on her memories.) To commit *tu quoque*, though, one would say that because the agent believes the inference rule is unjustified, or sometimes acts as if it were unjustified, the agent's conclusions gotten through the use of that inference rule must be *false or dismissable*. The epistemic principles above, which underwrite the supposition that there may be Type Two agent-relative fallacies, claim only that the agent's second-order beliefs about justification can defeat the agent's *epistemic justification* for certain first-order beliefs. Very possibly, they would *not*

defeat the epistemic justification for someone who lacks the relevant second-order beliefs.

Perhaps we should reject (EDJ) and (EDU), however. Reliabilist theories (which say, in their crudest form, that knowledge is true belief generated by a reliable process and that justified belief is any belief generated by a reliable process) are thought to counter the intuition behind such principles as (EDJ) and (EDU). So perhaps to get the verdict that one in the van Fraassen situation reasons fallaciously, one must adopt some sort of evidentialism or internalism about epistemic justification, and reject reliabilism. But it isn't so simple. Reliabilism has problems in characterizing processes. Is abduction the process van Fraassen employs in his daily life, drawing conclusions about student plagiarism and many other things? Yes, but so is 'trusting a source when one has evidence it is untrustworthy' or 'dismissing the testimony of an expert epistemologist on the subject of epistemology' and others, which are unreliable processes. (I am indebted to Richard Feldman (2005, *passim*) here.)

Additionally, if a reliabilist theory includes a "no defeater" condition, as Alvin Goldman's in fact does, then having evidence that abduction is unreliable can make one's abductive inferences unjustified, whereas one who had never given abduction any thought at all, would be justified in her abductive inferences. (Perhaps this is another case of epistemology destroying knowledge.) So it is unclear exactly what the verdict of the major epistemic theories would be for a case like this one. There is no clear reliabilist road to denying the possibility of Type Two agent-relative fallacies (by way of denying (EDJ) and (EDU), as various forms of reliabilism allow that one's evidence about one's evidence can affect the epistemic status of one's first-order judgments.

4. Conclusion

In this paper, I've explained the notion of an agent-relative fallacy and I've defended their plausibility. The possibility of such fallacies does not depend on the truth of any outrageous claims. In Type One cases, the thesis that the fallaciousness of an agent's inference is an extrinsic feature of the agent is dependent principally on the thesis that what an agent is thinking is an extrinsic feature of the agent (as per content externalism). In Type Two cases a particular inference is fallacious for one agent but not for another because the inference is epistemically justified for one agent, but not for the other. All we need here are plausible – even to reliabilists – epistemic descent principles about the possession

of epistemic defeaters.

NOTES

i I am grateful to Michael Veber, and to many members of the audience from my presentation on 1 July, 2010, at the Seventh Meeting of the International Society for the Study of Argumentation, in Amsterdam, for very helpful comments on an earlier version of this paper.

ii An epistemic ascent principle, on the other hand, moves from first-order epistemic claims to second-order ones. The so-called "KK" principle – if S knows that p, then S knows that S knows that p – is probably the best-known example of an epistemic ascent principle.

REFERENCES

Burge, T. (1979). Individualism and the mental. In P. French, T. Uehling & H. Wettstein (Eds.), *Midwest Studies in Philosophy Volume IV* (pp. 73-122). University of Minnesota Press: Minneapolis.

Feldman, R. (2005). Respecting the evidence. *Philosophical Perspectives, 19*, 95-119.

Kaplan, D. (1989). Afterthoughts. In J. Almog, J. Perry & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 567-614). Oxford University Press: Oxford.

Putnam, H. (1975). The meaning of "meaning". In *Mind, Language and Reality: Philosophical Papers: Volume 2* (215-271). Cambridge University Press: Cambridge.

Van Fraasen, B. (1989). *Laws and Symmetry*. Oxford University Press: Oxford.