volume III ♦ issue 1

2012

Logos & Episteme

an international journal
of epistemology

Romanian Academy
Iasi Branch

♦♦♦

“Gheorghe Zane” Institute
for Economic and Social
Research

INSTITUTUL EUROPEAN

Contact address:

Institutul de Cercetări
Economice şi Sociale „Gh.Zane"

Iaşi, str.T.Codrescu, nr.2, cod 700481
Tel/Fax: 004 0332 408922
Email: logosandepisteme@yahoo.com

www.logos-and-episteme.proiectsbc.ro

Logos & Episteme

ISSN-L: 2069-0533

ISSN: 2069-0533

The content of these materials does not necessarily represent the official position of the European Union and the Romanian Government.

For detailed information about the other programs co-financed by the European Union, we invite you to visit www.fonduri-ue.ro

This journal was edited within The Knowledge Based Society Project supported by the Sectoral Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number POSDRU ID 56815

TABLE OF CONTENTS

Articles

Debate

ARTICLES

KNOWLEDGE, PRACTICAL REASONING AND ACTION

Peter BAUMANN

ABSTRACT: Is knowledge necessary or sufficient or both necessary and sufficient for acceptable practical reasoning and rational action? Several authors (e.g., Williamson, Hawthorne, and Stanley) have recently argued that the answer to these questions is positive. In this paper I present several objections against this view (both in its basic form as well in more developed forms). I also offer a sketch of an alternative view: What matters for the acceptability of practical reasoning in at least many cases (and in all the cases discussed by the defenders of a strong link between knowledge and practical reasoning) is not so much knowledge but expected utility.

KEYWORDS: knowledge, practical reasoning, expected utility

Several authors have recently argued that there is a close connection between knowledge and practical reasoning. Williamson for instance says: "One knows *q* iff *q* is an appropriate premise for one's practical reasoning."[1] We can call this the "knowledge norm of practical reasoning"; for the sake of brevity we can refer to the claim as "the connection thesis." Stanley states that "one should act only on what one knows."[2] Hawthorne agrees that knowledge of a proposition is necessary for using it as a premise in acceptable practical reasoning[3]; he adds that it is both necessary and sufficient.[4] These kinds of claims are usually introduced by their defenders as intuitively plausible principles, supported by considerations and discussions of examples and cases,[5] like, for instance, cases involving lotteries

[1] Timothy Williamson, "Contextualism, Subject-Sensitive Invariantism, and Knowledge of Knowledge," *Philosophical Quarterly* 55 (2005): 231; see also Timothy Williamson, *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000), 47.

[2] Jason Stanley, *Knowledge and Practical Interests* (Oxford: Clarendon Press, 2005), 9.

[3] See John Hawthorne, *Knowledge and Lotteries* (Oxford: Clarendon Press, 2004), 29, 85, 174-175.

[4] See Hawthorne, *Knowledge and Lotteries*, 30; see also Jonh Hawthorne, Jason Stanley, "Knowledge and Action," *The Journal of Philosophy* 105 (2008): 571-590 and Jeremy Fantl, Matthew McGrath, *Knowledge in an Uncertain World* (Oxford: Oxford University Press, 2009). The thesis can easily be extended to multi-premise reasoning; the premises would, according to the connection-thesis, count as acceptable for one's practical reasoning as long as they are all known by the subject.

[5] See, e.g., Hawthorne, *Knowledge and Lotteries*, 85, passim, and Stanley, *Knowledge and Practical Interests*, 9-10, passim.

(especially in the case of Hawthorne). The recent debate on this topic as a whole (see below) has also been very much driven by such considerations of plausibility. More systematic theoretical motivations for the connection thesis have been rare and if they play a role at all they are rather working in the background: Hawthorne's and Stanley's subject-sensitive invariantism[6] goes very well with the connection thesis; for Williamson[7] the connection thesis is part of a *knowledge*-centered systematic epistemological theory. Apart from that, general dissatisfaction with expected utility accounts of practical reasoning (see below) might play some role in the background, too.[8]

There are several problems with such claims about knowledge and practical reasoning – whether knowledge is deemed necessary or sufficient or both necessary and sufficient for acceptable practical reasoning and action.[9] I will develop my objections step by step; most of them concern the necessity claim. I should stress from the beginning that I will follow the current debate on this topic and take "practical reasoning" to refer to instrumental reasoning here (if not indicated otherwise).

1. *The Main Example.* The example most often used in support of the connection thesis has to do with lotteries.[10] Hawthorne gives the example of

[6] See Hawthorne, *Knowledge and Lotteries* and Stanley, *Knowledge and Practical Interests*.

[7] See Williamson, *Knowledge and Its Limits*.

[8] Some might want to argue that being in a position to know *p* is necessary and sufficient for using *p* as a premise in acceptable practical reasoning. Alternatively, one might want to claim that being justified in believing *p* (or holding a justified true belief in *p*?) is the relevant condition. One would have to see the specific arguments for such claims; these arguments will be sufficiently different from the ones presented for the connection thesis. Hence, we should leave such related claims aside here.

[9] To be sure, acceptable practical reasoning has to meet further conditions (see John Hawthorne, Jason Stanley, "Knowledge and Action," *The Journal of Philosophy* 105 (2008): 578). This might make the search for sufficient conditions more difficult than the search for necessary conditions. For the sake of simplicity and because nothing hinges on it here, I will disregard concerns with these additional conditions. – It is very plausible to say (see Hawthorne, Stanley, "Knowledge and Action," 572) that if a person ought to have known a certain proposition and taken it into account in her practical reasoning but is in fact ignorant of its truth and thus does not take it into account, then her practical reasoning is defect; however, this does not show that it is the knowledge of the proposition she is missing here rather than, say, the justified belief in it; apart from that, there are, of course, also many cases where the subject is ignorant of the truth of a proposition without there being any "obligation" to know its truth. For the sake of simplicity, I will leave cases of justified ignorance aside here.

[10] Hawthorne, Stanley, "Knowledge and Action," 571-574 present several further cases in favour of their view. I won't go into them because they don't add anything new here.

someone who's been offered a penny for his lottery ticket. The person reasons as follows:

> "The ticket is a loser.
>
> So if I keep the ticket I will get nothing.
>
> But if I sell the ticket I will get a penny.
>
> So I'd better sell the ticket."[11]

Hawthorne claims that it is "clear enough" that this is a piece of bad reasoning and that "ordinary folk" agree with that. They would also agree that it is bad reasoning because the first premise is not known.[12]

First of all, it is not obvious that one can under no circumstances know a "lottery proposition" like the first premise of the quoted piece of practical reasoning.[13] The problem Hawthorne's book[14] is dedicated to[15] shows why this is not so clear: If I know that I will never be rich (which seems plausible) and if I also know that this entails that I won't win the lottery, then how could I not know that I won't win the lottery? I don't want to go into this difficult problem here but only point out that it is no trivial claim at all that one cannot know a lottery proposition. It is also not obvious – as we will see in more detail below (section 5) – which pieces of reasoning are good or bad and why.[16]

[11] Hawthorne, *Knowledge and Lotteries*, 29; see also 85.

[12] See Hawthorne, *Knowledge and Lotteries*, 29-30; see also Hawthorne, Stanley, "Knowledge and Action," 571-572.

[13] See Stephen Hetherington, *Good Knowledge, Bad Knowledge. On Two Dogmas of Epistemology* (Oxford: Clarendon Press, 2001), 102-107; Peter Baumann, "Lotteries and Contexts," *Erkenntnis* 61 (2004): 415-428.

[14] *Knowledge and Lotteries*.

[15] See originally Gilbert Harman, *Thought* (Princeton: Princeton University Press, 1973), 161.

[16] See also Rhys McKinnon, "Lotteries, Knowledge, and Practical Reasoning," *Logos & Episteme* 2 (2011): 225-231 who points out that Hawthorne and others need to make sure that the reasoning in a lottery case like the one above is bad not just because it violates principles of expected utility (see below). – What makes people say that the kind of reasoning discussed by Hawthorne, Stanley and others is bad? Here are two weaker but not unpopular arguments which I want to discuss briefly just to get them out of the way. First, someone might imagine that the person first bought a ticket in the expectation that it might win and then decided to sell it in the expectation that it will lose. In this case, the person – if she has not simply changed her mind – has incoherent expectations. Incoherence would then be the problem but not the lack of knowledge of a premise of the practical reasoning – the connection thesis does not seem relevant here. Apart from that, this kind of incoherence would not show that the practical reasoning leading to the selling of the ticket itself is bad; the incoherence would rather be

Let me begin with some rather straightforward objections against the connection thesis (sections 2-4) before discussing it in the light of an alternative view (sections 5-7).

2. *Gettierization.* Let us start with a point which is important enough to stress right at the beginning. If knowledge of a proposition is necessary for its use in acceptable practical reasoning, then gettierized subjects cannot engage in acceptable practical reasoning. But this does not seem plausible at all.[17] Compare Bo and Ben. Bo sees herself confronted with a raging bull. She comes to know there is a bull, engages in some swift practical reasoning and decides to run. Ben, however, finds himself, unknowingly, in raging bull facade county (where the locals frighten strangers with their exquisite raging bull facades). As it happens, he really is confronted with a raging bull. Given the circumstances, he doesn't come to know that.[18] Despite lack of knowledge, he engages in some swift practical reasoning and

between two different pieces of practical reasoning (the one leading to the buying of the ticket and the other leading to the selling of it).

Second, in discussion philosophers supporting the connection thesis often ask "So, if you decide to sell it – why did you buy it in the first place?" If this is not the incoherence charge again, then it can be taken either as the charge that there is no good reason to buy a lottery ticket in the first place ("and are you not admitting this yourself by deciding to sell it?"); or it can be taken as the charge that there is no good reason to sell it, once bought ("and are you not admitting this yourself by having bought it?"). On the first charge: There are many reasons to buy lottery tickets. If one can get emotionally involved while watching a movie knowing that it is all fictional, then why not get excited in a similar way about the slim prospects of winning a lottery? People have all kinds of reasons for buying lottery tickets and these reasons can be bad but they need not be. Furthermore, people's reasons and the quality of their reasons need not have anything to do with whether they know that their ticket will win or lose. On the second charge: Why not sell a lottery ticket? Some people do that even as their profession (whether or not they buy them in the first place) and there seems nothing wrong with it as such. Or is it specifically because one does not know that it will lose? This question leads us back to the main issue and requires more than just the raising of a rhetorical question ("So, if you decide to sell it – why did you buy it in the first place?").

[17] See also E.J. Coffman, *Old Dog Does New Tricks*, Ms., 2007; Jessica Brown, "Subject-Sensitive Invariantism and the Knowledge Norm for Practical Reasoning," *Noûs* 42 (2008): 167-189, sec.5; Jonathan L. Kvanvig, *The Value of Knowledge and the Pursuit of Understanding* (Cambridge: Cambridge University Press, 2003), 22; Mikkel Gerken, "Warrant and Action," *Synthese* 178 (2011): 535-6.

[18] See Alvin I. Goldman, "Discrimination and Perceptual Knowledge," in his *Liaisons. Philosophy Meets the Cognitive and Social Sciences* (Cambridge & London: MIT Press, 1992), 86 for this kind of example as well as Clayton Littlejohn, "Must We Act Only on What We Know?" *The Journal of Philosophy* 106 (2009): 464-5, 469, and Ram Neta, "Treating Something as a Reason for Action," *Noûs* 43 (2009): 687-688 who use it in a similar way than we do here; the argument here can be easily applied to all kinds of Gettier-cases.

decides to run. It seems implausible to assume that Bo's but not Ben's practical reasoning is fine because Bo but not Ben knows that there is a raging bull in the vicinity.

Gettierized belief in the narrow, original sense of the term[19] is justified true belief which does not amount to knowledge. Given the great variety of Gettier-like examples and the controversial nature of the notion of justification it might be better to use a broader notion of gettierized belief. In that sense, a belief is gettierized just in case it is true but does not constitute knowledge and this by no epistemic mistake of the subject. Epistemically blameless true belief which does not amount to knowledge is gettierized belief (in the broad sense). This explanation suggests that there need not be anything wrong with the practical reasoning of such a gettierized subject: If the subject fails to know the relevant proposition by no epistemic mistake of their own, then why should we blame them for their practical reasoning which is based on the relevant proposition? Sure, one would want to point out to the subject that she did not know what she seemed to know or that what seemed to be true is not in fact true; however, there is no reason to take this as a criticism of the subject's reasoning (how could she have done better?). Knowledge thus does not seem to be necessary for acceptable practical reasoning. One might want to object and argue that in such cases the subject is really violating a rule, though blamelessly. Lack of justified blame and the presence of good excuses, one could point out, do not entail that the rule of reasoning has not been violated. Since similar points will come up in the next two sections, I will reply to this kind of objection at the end of section 4.[20]

3. *Truth.* Is even the truth of a proposition necessary for the acceptability of using it in practical reasoning? Consider the following two cases. Jill is close to

[19] See Edmund L. Gettier, "Is Justified True Belief Knowledge?" *Analysis* 23 (1963): 121-123.

[20] One might also hold that the Gettier-objection only works against the connection thesis if one assumes in addition that the concept of knowledge can be reductively defined in terms of individually necessary and jointly sufficient conditions; according to this idea, it does not work against someone like Williamson who holds that the concept of knowledge cannot be reductively defined (see Williamson, *Knowledge and Its Limits*). I disagree. While it is true that Gettier cases are a problem only for reductive definitions of knowledge, it is also true that they pose a problem for any defender of the connection thesis. Defenders of reductive as well as of non-reductive accounts of knowledge can easily agree that there are gettierized subjects (that is, subjects who meet certain conditions while lacking knowledge); thus, the question whether such subjects are entitled to use the believed proposition as a premise for practical reasoning even if they don't know it makes a lot of sense for both parties. Williamson, for instance, agrees, too, that there are Gettier cases; he thus needs to answer the Gettier-objection above like everyone else.

dying of thirst when she finds some water. She comes to know it is water and engages in some quick practical deliberation:

> This is water.
>
> If I drink this, I will survive.
>
> If I don't drink this, I will die.
>
> So, I better drink this.

She drinks the water and survives. Jack, however, has recently travelled to twin earth but doesn't know about the differences between earth and twin earth.[21] He finds himself in a situation identical to Jill's as far as the subject’s perspective is concerned. However, what he takes to be water is really something different, namely t-water. Fortunately, t-water is as good for survival as water. Jack engages in some quick practical deliberation:

> This is water.
>
> If I drink this, I will survive.
>
> If I don't drink this, I will die.
>
> So, I better drink this.

He drinks the t-water and survives. Should we really say that there is something wrong with Jack's practical reasoning but not with Jill's because Jill knows that there is water whereas Jack's belief that there is water is not even true? This does not seem plausible.[22] True belief does not seem necessary for acceptable practical reasoning based on that belief; it follows that knowledge is also not necessary. Perhaps Hume was right after all and belief (and desire) is all we need for an account of good practical reasoning? Again, one might object that all this only shows that the subject is blameless and excused but not that no rule of practical reasoning has been violated. I will get back to this kind of point at the end of the next section.

4. *KK.* What if S knows that *p* but does not know that he knows that *p*? Suppose S is unsure and does not know whether he knows that *p* (see Radford 1966). Assuming that this is compatible with S's knowledge that *p*, S could have a

[21] See Hilary Putnam, “The Meaning of ‘Meaning,’” in his *Mind, Language and Reality. Philosophical Papers, vol. 2* (Cambridge: Cambridge University Press, 1975), 215-271; if one does not like the semantic externalism involved in the example one can easily modify it by replacing reference to t-water by reference to some liquid indistinguishable from water.

[22] See also Trent Dougherty, *Knowledge and Context-Sensitive Norms: A Defense of Simple Moderate Invariantism*, Ms., 2007.

good reason not to simply act on the proposition that *p*. Sure, if "not acting" is not a serious alternative, that is, if not acting on any proposition relevant to the practical issue at hand were to risk bringing about very bad consequences, then S ought to act on some proposition, and if the proposition that *p* seems more likely to be true than any alternative proposition, then S ought to act on the proposition that *p*. However, this might not be the case: "Not acting" might be unproblematic or better than acting on a proposition that "might," according to S's worries, turn out to be false; apart from that, there could also be competing propositions which seem to S more likely to be true. In such cases, the subject does not have a good reason to act on the proposition that *p* and even has a good reason not to act on the proposition that *p*. Paul might know that the answer to the 1 Million Pound question is "Teheran." However, he is cautious and really not sure whether he knows that: Didn't he make bad mistakes about geography before? Suppose Paul can ask a friend about this. In such circumstances he should not simply go ahead and give "Teheran" as his final answer. Knowledge of a proposition is thus not sufficient for its use in good practical reasoning (see fn.1 above). Would knowledge that one knows be? Perhaps – but that thesis is much stronger and much less interesting.[23]

What about the reverse case where someone (by no mistake of their own) does not know that *p* but has very good reasons to think they know that *p*? One can think of cases of gettierization here or of cases like the above one about water. Isn't it acceptable then to act on a proposition which is not known (whether that proposition is true or false)? Even if the practical stakes of getting it right are very high – which would be relevant to the case of an unknown false proposition), the subject ought not go by anything but their best reasons; if these reasons suggest that *p*, then S ought to act on *p*. If that is correct, then knowledge is not necessary for good practical reasoning.

Hawthorne and Stanley discuss such cases[24] and propose that they are not counter-examples to the connection thesis (or what they call the "knowledge-rule") but rather cases where the subject has a good excuse and thus deserves no blame for violating the rule. This kind of reply can also be used against the cases mentioned in the last two sections.

[23] Williamson, "Contextualism, Subject-Sensitive Invariantism," sec.V briefly discusses the appropriateness of practical reasoning from a known premise when one does not know that one knows the relevant proposition; however, what he proposes here (appropriateness depending on and varying with subject's stakes) does not solve the problem above for the straightforward connection thesis. I will therefore not further go into this here. See also a brief remark in Jonathan L. Kvanvig, "Against Pragmatic Encroachment," *Logos & Episteme* 2 (2011): 80-1.

[24] Hawthorne, Stanley, "Knowledge and Action," 573, 586; see also Neta, "Treating Something," 688-9.

However, there is a general methodological problem with this kind of reply: Can't one always say that? Or at least way too often? Hawthorne and Stanley don't make it clear how one can distinguish between a counter-example to the knowledge-rule and the case of an excused violation of the rule. Hawthorne and Stanley use this kind of reply – the interpretation of what might at first sight look like a counter-example as a mere excusable violation of their rule – quite often.[25] Without further arguments to the contrary, one is entitled to at least let it cut both ways: While Hawthorne and Stanley might say that something is an excusable violation of their rule, the sceptic about the connection thesis seems to have at least as much reason to see it as a counter-example. I don't want to put too much weight on this point and rather leave it at that; however, this should already raise some doubts about the connection thesis.[26]

But wouldn't we, after discovering that our belief in some premise was gettierized or that we were mistaken about what it is that we refer to or that we were wrong about our epistemic states (see this and the last two sections), criticize our reasoning as inadequate? And wouldn't that show that knowledge is required for appropriate practical reasoning after all? I don't think so. Even if one acknowledged this kind of "drawback" one wouldn't have to concede and should in fact not concede that there was something wrong with one's original reasoning: there wasn't anything wrong with that. The gettierized person might find herself lucky when noticing that she had been gettierized but she would not have any reason to see her practical reasoning itself as deficient. Sure, she might concede that the basis of her reasoning was not as good and solid as she thought; but even then she would be right to insist that there was nothing wrong with her reasoning as such, given that there was nothing she could have done to improve the basis of her reasoning. Similar things hold for subjects who are mistaken about referents or about their epistemic state (see above). Even if the premises on which her reasoning were based turned out to be false, she ought not to accept blame for her reasoning itself.

5. *An Alternative: Expected Utility.* Consider again the original piece of practical reasoning:

(Case 1)

The ticket is a loser.

[25] Hawthorne, Stanley, "Knowledge and Action," 585-586 also remark that our intuitions about what constitutes a counter-example and what not are unclear in cases of failure of luminosity where one is in a particular (mental) state without knowing that one is.

[26] For more on excuses see Gerken, "Warrant and Action," 537-544.

So if I keep the ticket I will get nothing.

But if I sell the ticket I will get a penny.

So I'd better sell the ticket. [27]

Why is it bad? Is it because the first premise in not known? Or is it for some other reason?

Compare (Case 1) with another case. John finds a lottery ticket with the number 666 on the street. He knows a crazy collector of items with the number 666 on them who will offer him $5,000 for the ticket. He reasons in the following way:

(Case 2)

The ticket is a loser.

So if I keep the ticket I will get nothing.

But if I sell the ticket I will get $ 5,000.

So I'd better sell the ticket.

What if, in addition, John did not find the ticket but was paid a good sum by the former owner of the ticket to take it because he thought that keeping a ticket with that number will bring terrible bad luck? Hawthorne remarks in a footnote[28] – without further argument – that it doesn't make a difference whether the ticket was free or not; I find it hard to see how this could not make a difference.

Finally, Hawthorne himself briefly mentions the case of a 10,000 ticket lottery with a $5,000 prize where one ticket costs a cent. It would be irrational "to decline [buying a ticket] on the basis of your 'knowledge' that the ticket will lose."[29] Here is the reasoning for this case:

(Case 3)

That ticket is a loser.

So if I buy the ticket I will get nothing.

But if I don’t buy the ticket I will keep one 1 cent.

So I'd better not buy the ticket.

(Case 1) and (Case3) strike us as cases of bad practical reasoning while (Case 2) is not so bad if not quite good. Why is that? All the cases share the same logical form.

[27] See Hawthorne, *Knowledge and Lotteries*, 29.

[28] See Hawthorne, *Knowledge and Lotteries*, 29, fn. 74.

[29] Hawthorne, *Knowledge and Lotteries*, 85.

We get closer to an answer if we make the relevant implicit assumptions about odds and stakes explicit. In (Case 1) and (Case 2) we can assume that the subject owns a ticket with a minute chance of winning (say 1 out of 1 Million) while the chances are much better in case (3) (1 out of 10 000). Let us assume the prize in (Case 1) and (Case 2) is 10 000 Dollars.

It is tempting to use Expected Utility Theory to analyse such cases and interpret them as cases of expected utility reasoning.. The basic idea is that in cases like the above ones appropriate practical reasoning identifies from the feasible set of available acts the act with the highest expected utility (whether it is explicitly guided by this idea or not). Given a finite and exhaustive set of mutually exclusive set of circumstances c_1 … c_n,[30] given further a utility function U of the subject which maps the outcome of each act in a given circumstance to a (measurable) utility u_1 … u_n, and given, finally, a probability function P of the subject which assigns each circumstance or outcome a certain probability p_1 … p_n, the expected utility EU of an act A can be characterized as the sum p_1 x u_1 + … p_n x $u_{n.}$ Or:

$$EU = \sum_i p_i \text{ x } u_i$$

Given EU, we cannot represent practical reasoning any more in the form used above, namely simply as deductive inferences from given premises to a conclusion. Otherwise we could not explain why the pieces of reasoning differ in quality, given that they share the same logical form. Rather, we have to take into account that the reasoner has different credences in the different propositions. This, together with the different utilities, explains why the reasoning is good in one case and bad in the two other cases above.

Let us apply this idea to our 3 cases. For the sake of simplicity we may assume that the scale of the subject’s utilities for money can be mapped by a positive linear transformation onto the scale of monetary values (nothing hinges on this simplification). Then the expected utility of keeping the ticket in (Case 1) is 1/100 expected Dollars; the expected utility of selling the ticket is also 1/100 expected Dollars. Both acts have the same expected utility and there is thus no reason (*ceteris paribus*) to prefer one to the other, given the theory. For this reason, the inference to the conclusion of (Case 1) is not a good one.

In (Case 2) the expected utility of keeping the ticket is, 1/100 expected Dollars while the expected utility of selling the ticket is 5000 expected Dollars. So, selling looks like the rational thing to do.

[30] If the set is infinite, things are more complicated; fortunately, the assumption of finite sets is unproblematic here and nothing hinges on it.

In (Case 3), finally, the expected utility of not buying a ticket is 0 Cents while the expected utility of buying a ticket is 49 expected Cents. Hence, it would be foolish not to buy a ticket.

According to the connection thesis, all 3 pieces of practical reasoning are bad because some premise is unknown. According to Expected Utility-Theory, the first and last case (1, 3) are cases of bad reasoning while the middle case (2) is a case of good reasoning. In cases like these, Expected Utility-Theory seems to give the correct answers while the connection thesis doesn't.[31]

Here is another case Hawthorne brings up:

> "I will be going to Blackpool next year.
>
> So I won't die beforehand.
>
> So I ought to wait until next year before buying life insurance."[32]

Given that the life insurance offers a reasonable deal, this is an example of bad reasoning. But, again, the reason it is bad is that it goes against the expected utilities and not the lack of knowledge of the premises of the inference (I leave it to the reader to go through the numbers for particular examples). Similarly, there are perfectly acceptable pieces of reasoning in favour of buying life insurance, even if the subject does not know the relevant premises of her reasoning. While the connection thesis excludes certain propositions from acceptable practical reasoning, namely the unknown ones, the principle of expected utility is much more liberal and accepts them all; the constraints of the latter view rather concern the question how one ought to use a proposition (namely according to Expected Utility-Theory) rather than whether to use it.[33]

6. Expected Utility versus Knowledge? But is it really true that practical reasoning guided by the idea of maximizing expected utility does not require

[31] I say "seems" because I have not offered a formal argument that Expected Utility-Theory always gives the right results in such cases. However, plausibility is enough for my purposes here, given that the main aim of this paper is to argue against a view and not so much for an alternative.

[32] Hawthorne, *Knowledge and Lotteries*, 175; see also Hawthorne, Stanley, "Knowledge and Action," 571.

[33] On expected-utility accounts as alternatives see also the brief remarks in Hawthorne, Stanley, "Knowledge and Action," 580-585, Richard Feldman, "Knowledge and Lotteries," *Philosophy and Phenomenological Research* 75 (2007): 222-225, Alan Goldman, "Knowledge, Explanation, and Lotteries," *Noûs* 42 (2008): 471-472, Igor Douven, "Knowledge and Practical Reasoning," *Dialectica* 62 (2008): 101-118, and Richard Fumerton, "Fencing Out Pragmatic Encroachment," *Philosophical Perspectives* 24 (2010): 247; for acceptable reasoning using unknown lottery propositions see also Littlejohn, "Must We Act," 471-472.

knowledge of its premises? Expected Utility-reasoning can also be characterized informally by the following schema (restricted here, for the sake of simplicity, to the case of 2 feasible acts):

> (1) I can do A_1 or A_2.
>
> (2) Given the possible circumstances and outcomes, A_1 has higher expected utility than A_2.
>
> (C) Hence, I should do A_1.[34]

Or, more briefly (and very roughly):

> Doing A_1 is the best means for the attainment of my ends.
>
> Hence, I should do A_1.

Does acceptable practical reasoning along such lines require that one knows one's feasible set of acts, the set of possible circumstances of action and the different outcomes of given acts in given circumstances (instead of having some belief about this which falls short of knowledge)? Since there are many cases where we engage in acceptable practical reasoning and since we often do not know these things, it seems very plausible to assume that the acceptability of practical reasoning does not require knowledge of these things. In other words, acceptable practical reasoning is compatible with lack of knowledge of at least some of its premises. I won't go more into lack of knowledge of one's options, circumstances and outcomes here because the point just made seems very plausible; it is uncontroversial (or not even an issue) in the debate on the connection thesis.

But what about probabilities and utilities? Doesn't acceptable practical reasoning require knowledge of at least them? We have seen that knowledge of the premises in cases like (Case1), (Case 2) and (Case 3) are not required. This is already an interesting result concerning the standard form in which practical inferences are usually presented by defenders of the connection thesis. But can't we restrict and reformulate the connection thesis and claim that acceptable practical reasoning though not requiring knowledge of all of its premises still does require knowledge of some premises, namely of those stating probabilities and utilities?

Let us take the case of utilities first. It is quite plausible, I think, to assume that agents often have mistaken views about or are ignorant of some of their utilities. However, even in such cases acceptable practical reasoning seems possible. Suppose I wonder whether I should spend the evening out with friends or rather

[34] Sometimes it is said that the conclusion of a practical inference is an action rather than a proposition; we can leave this complication aside here.

alone at home.[35] I reason in the light of what I take to be best for me. Suppose that I assume on the basis of good evidence (about myself) but falsely that it would be better for me to stay home alone tonight. Still, my reasoning that I should stay home alone strikes me as good while any reasoning resulting (on the same basis) in the conclusion that I should go out would appear unmotivated and foolish even if this option would really be better for me. To know one's own utilities is certainly an important advantage but it seems forced to say that it is also a necessary condition for acceptable practical reasoning.

I don't want to pursue the issue of knowledge of one's utilities any futher here but rather, finally, go into knowledge of probabilities. This is a topic defenders of the connection thesis have commented upon – while they haven't said much if anything about knowledge of utilities.

Stanley points out that even if one is dealing with probabilities (or expected utilities) in one's practical reasoning, one still needs knowledge, namely knowledge of those probabilities[36]). Take an agent who's deliberating about the question whether they should buy a ticket in a 10 ticket lottery with a $10,000 prize for just one cent. If this reasoning is acceptable, then the agent knows the relevant probabilities, so the idea.

But does she really have to get the probabilities right in order for her reasoning to count as a good practical reasoning? Suppose both Ann and Barbara have received exactly the same information about some lotteries, – Ann about a lottery she is considering and Barbara about a different lottery she is considering. Suppose further that they both have no reason to be suspicious about the information. The only difference is that in Ann's case the information is correct whereas in Barbara's case it is incorrect. Should we really be so "externalist" about practical reasoning as to say that Ann's but not Barbara's practical reasoning is good because only Ann but not Barbara knows the probabilities? This defence of the connection thesis against the expected utility objection comes at a high prize: One would have to accept a very controversial theory of practical reasoning, namely externalism (the idea that the quality of practical reasoning is, at least partly, determined by factors which might not be accessible to the subject; defenders of the connection thesis have not

[35] See, e.g., Friedrich Waismann, "Ethics and the Will," in Friedrich Waismann, Josef Schächter, Moritz Schlick, *Ethics and the Will. Essays*, eds. Brian McGuinness and Joachim Schulte (Dordrecht: Kluwer, 1994), 53-137, sec.17 for the issue and similar examples.

[36] See Stanley, *Knowledge and Practical Interests*, 10, Jason Stanley, "Replies to Gilbert Harman, Ram Neta, and Stephen Schiffer," *Philosophy and Phenomenological Research* 75 (2007): 203-206, and Hawthorne, Stanley, "Knowledge and Action," 581-585.

offered any argument to this conclusion).[37] Apart from that, the theory would demand a lot from deliberators, namely knowledge about (objective) probabilities. But are probabilities "out there" for us to know? The defence discussed here thus comes with a lot of substantial commitments on other topics: One has to accept not only a strong externalism about practical reasoning but also a particular, controversial view of probability. This in itself speaks against this defence manoeuvre.

Perhaps those who want to make this kind of move should then rather think of probabilities as subjective or as epistemic (this is Hawthorne's and Stanley's move[38]). However, if one does that, it becomes unclear why the agent has to know their probabilities and why it is not sufficient just to rely on them (e.g., as her betting dispositions). Her behaviour would then just express his probabilities (together with her utilities). Furthermore, in this case the move towards probabilities seems besides the point. Knowledge of the probabilities wouldn't be knowledge of external facts, about the world rather than the agent; however, knowledge of the world (and not of one's own mind) would be what is needed in practical reasoning if knowledge of any interesting kind is needed at all.

Hawthorne and Stanley offers a reply to this objection.[39] Suppose my epistemic probability that the restaurant is on the left is .6 because 3 out of 5 persons I asked told me so. Then I can deliberate and act on the known proposition that 3 out of 5 persons told me it's on the left, and I don't need to act on the known epistemic probabilities. It is very doubtful whether this reply helps Hawthorne and Stanley. Why is my knowledge that 3 out of 5 persons told me the restaurant is on the left relevant to the question whether I should go right or left – if not because it implies something about or simply reduces to a belief about the relevant probabilities? So, we're back with knowledge of probabilities.

To summarize: If the defenders of the connection thesis refer to objective probabilities, then they run into serious problems (see above); if they only talk about knowledge of subjective or epistemic probabilities, then the connection thesis loses most of its bite: The knowledge necessary or sufficient for good practical reasoning would be about subjective or epistemic states and not about facts in the external world – which is what the knowledge relevant to the connection thesis should be about if the thesis is supposed to be interesting.[40]

[37] To be sure, it is "better" to know the probabilities than not to know but this does not entail that there is something wrong with the reasoning of the unknowing subject.

[38] In Hawthorne, Stanley, "Knowledge and Action," 584-585.

[39] Hawthorne, Stanley, "Knowledge and Action," 584-585.

[40] See Stephen Schiffer, "Interest-Relative Invariantism," *Philosophy and Phenomenological Research* 75 (2007): 188-195, sec.1; Schiffer also points out that the subject need not have any concepts of probability.

The upshot of all this is that appropriate practical reasoning does not require knowledge of the premises of inferences like the ones in (Case 1), (Case 2) or (Case 3). It also turned out that the usual representation of practical reasoning in these kinds of cases is a bit elliptical and needs to be reformulated, namely, as we say, in terms of expected utilities. An alternative view of practical reasoning as based on the idea of expected utility accounts better for our judgments about good and bad practical reasoning. This view also allows for acceptable practical reasoning without knowledge of the premises. (We should always remember that "practical reasoning" refers to instrumental reasoning here).

Sure, no practical reasoner is ever completely wrong or ignorant about all the facts relevant to her reasoning. And perhaps some knowledge of some propositions is required for practical reasoning. For instance, a practical reasoner who does not know some basic facts about how actions intervene in the worlds cannot count as a good practical reasoner. Conceding this is, however, far from agreeing with the much stronger connection thesis or the claim that good practical reasoning requires knowledge by the subject of all the propositions used as premises in their reasoning.

I have focused here very much on expected utility and there are several objections one might raise against this view. There are other, non-maximizing, conceptions of practical reasoning. "Satisficing" views, for instance, hold that enough is enough and that one need not or should not maximize goods.[41] However, similar points can be made, *mutatis mutandis*, on the basis of these kinds of views. They also don't entail or support the connection thesis. I won't go into any detail here.

There are also choice situations where the subject has no idea concerning the relevant probabilities. In such cases, Expected Utility-Theory is inapplicable. It is not clear[42] whether there are any principles or rules of good practical reasoning for such cases and, if so, which ones (*Maximin*?). This kind of case, however, does not seem to help the defender of the connection thesis much, given that subjects know even less here than in the cases considered above. What about the other extreme, namely cases where the subject is certain about the outcomes of given acts in given circumstances? These cases also don't give the connection thesis any advantage over Expected-Utility-theory; the only relevant difference here is that the probabilities for the circumstances go up to 1 in one case and down to 0 in all other cases. It is interesting to notice that the defenders of the connection thesis typically present as

[41] See Herbert Simon, *Reason in Human Affairs* (Oxford: Blackwell, 1983), Michael A. Slote, *Beyond Optimizing: A Study of Rational Choice* (Cambridge: Harvard University Press, 1989), Barry Schwartz, *The Paradox of Choice: Why More Is Less* (New York: ECCO, 2004).

[42] See the overview in Michael Resnik, *Choices. An Introduction to Decision Theory* (Minneapolis & London: University of Minnesota Press, 1987), ch. 2.

cases of bad reasoning cases where the subject has some idea about the probabilities. The argument above, if it stands, seems sufficient against the connection thesis.

There are other principles or rules of good practical reasoning which often converge with Expected Utility-principles, like, e.g., the principle of dominance.[43] An act A dominates another act B just in case under no circumstance is the outcome of A worse than the outcome of B and under some circumstance it is better. The principle of dominance says that one should choose the acts which dominate all other feasible acts in a given situation of choice. Obviously, this principle is compatible with the lack of knowledge of all premises of the practical inference. There is no need to go into the details here. I will rather continue by illustrating the advantage of Expected Utility-theory over the connection thesis a bit more.

7. *Additional Considerations.* I would like to end by adding some further considerations against the connection thesis. The main weight lies on what has been said above but what follows should also be taken into account. I will discuss extreme stakes (a), further explanatory advantages of expected utility accounts (b) and end with some brief remarks on implications for epistemological scepticism and moral cognitivism (c).

(a) Considerations of extreme lotteries (or, more generally: of situations where very much is at stake) add further reasons to doubt the connection thesis. It is worth going into it briefly. Again, this rather supports the idea that it is expected utility and not knowledge that matters for practical reasoning.[44] Consider a 100 billion ticket lottery with a $100,000 prize; you got your ticket for free but have the chance of selling it for $90,000. It would be foolish not to sell the ticket even if you don't know that it won't win. This suggests that knowledge of a proposition is not necessary for acceptable practical reasoning based on that proposition.[45] Or take something you know for certain, like "if it rains, then it rains." It would still be foolish to bet your life on it, at least in normal circumstances (if you think that one

[43] As is well known, plausible principles of practical reasoning can conflict with each other. Newcomb's paradox is one very well-known case where the principle of dominance conflicts (or seems to conflict) with a principle of maximizing expected utility (see, e.g., Robert Nozick, "Newcomb's Problem and Two Principles of Choice," in his *Socratic Puzzles* (Cambridge: Harvard University Press, 1997), 45-73.

[44] See Stewart Cohen, "Knowledge, Assertion, and Practical Reasoning," *Philosophical Issues* 14 (2004), 487; Hawthorne, *Knowledge and Lotteries*, 148; Brian Weatherson, "Can We Do without Pragmatic Encroachment?" *Philosophical Perspectives* 19 (2005): 438-440; Dougherty, *Knowledge and Context-Sensitive Norms*; Schiffer "Interest-Relative Invariantism," sec.1; Brown, "Subject-Sensitive Invariantism," sec.7; Janet Levin, "Assertion, Practical Reasoning, and Pragmatic Theories of Knowledge," *Philosophy and Phenomenological Research* 76 (2008): 377-380.

[45] See also Levin, "Assertion, Practical Reasoning," 377-380.

should assume probability 1 for logical tautologies and should bet everything in such cases of maximal certainty, then you should modify the example and consider a case where the probabilities are extremely high but still below 1). This suggests that knowledge of a proposition is also not sufficient for the legitimate use of it in practical reasoning.[46] Expected Utility accounts, in contrast, have no problem at all accounting for such cases.

Hawthorne shortly mentions the last point and makes the following remark, sketching a response: "... we should consider whether knowledge of any proposition can be destroyed by environments in which a suitable bet is offered. One option, of course, is to think that the sketched connection between knowledge and practical reasoning is only roughly correct."[47] Let us leave the latter option ("One option …") aside here: One would have to spell out in detail in what ways this is only roughly correct and how it could be modified; since this has not been done yet, this idea is hard to discuss. Consider rather the first idea: If the stakes are high enough, then the person does not know the relevant proposition (e.g., that if it rains, it rains). How good is that reply? Well, it will be attractive to those who like Hawthorne or Stanley (but not Williamson) hold that knowledge depends on what is at stake for the subject. It won't cut much ice for those who don't accept that theory.[48] So, the dialectical weight of this rejoinder is limited.[49] Apart from that, it is hard anyway to imagine circumstances in which "If it rains, then it rains" would become unknown.[50]

[46] See for a different argument here: Thomas M. Crisp, "Hawthorne on Knowledge and Practical Reasoning," *Analysis* 65 (2005): 138-140, Jeremy Fantl, Matthew McGrath, "On Pragmatic Encroachment in Epistemology," *Philosophy and Phenomenological Research* 75 (2007): 558-589, and the remarks in Keith DeRose, *The Case for Contextualism. Knowledge, Skepticism, and Context, vol.1* (Oxford: Clarendon Press, 2009), 252-254, 262-268; also see fn. 4 above.

[47] Hawthorne, *Knowledge and Lotteries*, 177, fn.37.

[48] See also Brown, "Subject-Sensitive Invariantism," sec.7.

[49] See also the discussion in Hawthorne, Stanley, "Knowledge and Action," 587-589 where the authors do not commit themselves to a particular strategy against the objection above.

[50] One further criticism of the idea that an agent might know that *p* but still not be entitled to act on what they know says that this would allow even a knowing agent to further check the evidence. However, to say something like "I know she'll be at the party but let me check!" sounds weird (see Hawthorne, *Knowledge and Lotteries*, 148-149). Does it really? (see Brown, "Subject-Sensitive Invariantism," sec.7, Jennifer Lackey, "Acting on Knowledge," *Philosophical Perspectives* 24 (2010): 361-382 but also Neta, "Treating Something," 697.) What about the surgeon who claims to know that he is supposed to take out the appendix but decides to check the file again, "just in case"? Even if it does sound weird to say something like "I know she'll be at the party but let me check!": It does not seem to be relevant here. In the case of "If it rains, then it rains" we might not even be able to think of further evidence one could check. And

(b) Compare two pieces of knowledge. Suppose you know some very complex proposition about elementary particles and you also know that you exist now. You would certainly bet much more on the latter than on the former. This in itself does not show that knowledge isn't necessary for practical reasoning but it suggests at least that something else, the subject's probabilities in connection with expected utility, is doing the explanatory work. An expected utilities view which takes probabilities into account can, in addition, explain something the connection thesis cannot explain: why one should bet more on the second than on the first proposition.

Compare a strong belief or conviction which doesn't amount to knowledge (let us assume "I won't win the lottery" is an example) with some piece of knowledge the subject is not nearly as certain of ("The Kopenhagen view on quantum mechanics is the right one"), assuming here that knowledge does not entail a probability of 1.[51] It would be foolish not to bet more on the former than on the latter. This, again, suggests that what matters for practical reasoning is the subject's probabilities or expected utility and not knowledge. An expected utilities account, again, has an explanatory advantage over the connection thesis here: It can explain why the subject should bet more on one proposition than on the other.

Finally, suppose I have to decide whether to buy or not to buy a lottery ticket. Suppose further that neither do I know that the ticket offered will lose nor, of course, do I know that it will win. On the basis of what am I going to make my decision? Apparently not on the basis of knowledge concerning winning or losing.[52] My reasoning should be rather based on probabilities. – Apart from all this, Expected Utility can easily explain cases of Gettierization, false belief, lacking knowledge that one knows and false but justified belief that one knows (see sections 2-4 above).

even if all the evidence has been checked, it would still not be appropriate to bet one's life on the trivial conditional.

[51] If one disagrees here, holding that knowledge that *p* entails a probability of 1 for *p*, then this example will not work. However, this kind of defense is costly: One has to make very controversial and not intuitively plausible assumptions about knowledge and probability in order to defend a thesis, the connection thesis, which is supposed to be intuitively plausible. It is in general not a good strategy to defend the (allegedly) plausible with reference to the controversial. See also Jeremy Fantl, Matthew McGrath, "Critical Study of John Hawthorne's *Knowledge and Lotteries* and Jason Stanley's *Knowledge and Practical Interests*," *Noûs* 43 (2009): 185.

[52] Hawthorne, Stanley, "Knowledge and Action," 581, fn.10, mentions this point shortly without discussing it.

(c) Consider a certain kind of epistemological scepticism which does not so much raise doubts about the existence of the external world but rather denies that anyone ever knows anything in particular about the world. Knowledge, according to this kind of scepticism, requires that we meet a condition (e.g., to be able to rule out that we are dreaming at the moment) which we cannot meet, given our actual constitution. Call this "scepticism." Now, either it is (metaphysically) possible or not possible that scepticism is true. Suppose the defender of the connection thesis accepts that there is a possible world in which scepticism is true. Perhaps we do know lots of things in the actual world but would fall short of the conditions for knowledge in some possible world. Would we then (in that possible world) never be entitled to practical reasoning concerning what to do in the world? This seems very implausible[53]; however, the defender of the connection thesis would, it seems, have to say exactly that – if he allows for the possibility of scepticism being true. The only alternative is to deny the latter and argue that scepticism is necessarily false. This, however, looks like a very strong thesis in need of much argument, and the connection thesis itself does not provide such an argument (neither does the general account of knowledge Hawthorne or Stanley favour). Hence, if the defender of the connection thesis does not want to go with the first, rather implausible option, they will have to accept a very substantial and controversial thesis concerning scepticism which is very much in need of argumentative support. This does, of course, not show that the connection thesis is false but it reduces its attractiveness drastically. One would first have to decide whether scepticism is necessarily false before one can reach a view about the connection thesis.

Another problem arises with respect to morality. Moral reasoning is an important type of practical reasoning. I will keep my remarks short here, also because the defenders of the connection thesis have said (next to) nothing about this aspect. Consider the following plausible piece of moral reasoning:

(1) That person is in need of my help

(2) If someone is in need of my help, then (given certain background conditions), I ought to (better) help that person

(3) Hence, I ought to (better) help that person.[54]

[53] See Dougherty, *Knowledge and Context-Sensitive Norms*, who argues that a subject in a sceptical scenario would have a justified false belief but her practical reasoning would remain unaffected; see also a brief passage in Kvanvig, "Against Pragmatic Encroachment," 81.

[54] The background conditions mentioned here are of the following sort: I can help easily, there are no strong reasons not to help that person, etc. We can disregard these complexities here.

If the connection thesis also covers moral reasoning and not just instrumental reasoning and if it requires knowledge not just of the factual premises but also of the normative ones, then the connection thesis implies that some form of moral cognitivism must be true: The reasoner in our example needs to know the normative premises (e.g., (2)), too. Normative premises are knowable and truth-apt. If only the moral expressivists had known about the relation between knowledge and practical reasoning! But can a thesis in epistemology really have substantial implications in meta-ethics like moral non-cognitivism? We have good reason to be sceptical here.

To conclude, one should not expect decisive arguments in the debate about the relation between knowledge, practical reasoning and action. The considerations offered here, however, make a strong case against the connection thesis. It remains to be seen whether the defenders of the connection thesis can come up with convincing replies.

THE SIGNIFICANCE OF COMBINING FIRST-PERSON AND THIRD-PERSON DATA IN NEUROSCIENCES: AN EXAMPLE OF GREAT CLINICAL RELEVANCE

Dana Maria BICHESCU-BURIAN[1]

ABSTRACT. Both perspectives, the one of the first and the one of the third person and their interrelation are necessary for the progress of consciousness research. This progress presupposes the systematic and productive collaboration between philosophy and neuroscience and cognitive science. While the philosophy of mind deals with working out clear conceptual implications and argumentative coherency in this area and critically follows the state of the art in this regard, the mission of neuro- and cognitive sciences is to develop and employ useful methods for the approach of the main problems of consciousness. I discuss this necessity by the example of research on implicit and explicit memory processes. Implicit and explicit memory processes are essential for the understanding and treating several psychological and neurological disorders. Among these, memory deficits play a crucial role in stress-related disorders, such as PTSD, dissociative disorders, and borderline personality disorders. Criticism has been exercised with regard to neglect of subjective experience in the research of memory processes, as well as the inadequate application of the concept of consciousness, usually leading to confusion. However, a step forward has already been taken in the research of memory processes. For example, the psychotraumatology research provided important advances in understanding the undelying distorsions in implicit and explicit memory procesess by employing combined assessments of both first-person and third-person data. Such multimodal research approaches delivered an exemplary model for the scientific investigation of mental processes and disorders and their neuronal substrates.

KEYWORDS: first-person data, third-person data, philosophy of mind, neurosciences, cognitive sciences, clinical psychology, implicit and explicit memory

1. An unsolved problem

How does consciousness arise in the physical world? This is a question that has preoccupied many scientists and philosophers for centuries. An important part of the problem, which corresponds to the project of empirical consciousness research,

[1] The paper was made within *The Knowledge Based Society Project* supported by the Sectorial Operational Program Human Resources Development (SOP HRD), financed by the European Social Fund, and by the Romanian Government (POSDRU ID 56815).

is to understand how a variety of subjective universes can constantly develop and fade away in our objective universe. The philosophical part of the problem is to understand how we ourselves can embody such subjective universes and, above all, what all that really means.[2]

The subdivision of the problem in the philosophy of consciousness

1. In philosophy, the main problem is the *ontological* one, which deals with *the nature of mental processes.*[3] Main issues in this area are: Can mental phenomena be attributed to physical phenomena? Can mental states be realized physically?

2. *Epistemologically* there is a distinction between:

a. *the problem of knowledge about our own mental states* and

b. *the knowledge about the mental states of other.*[4]

The first problem is called the problem of *privileged access to one's own mental states* or the problem of *first-person perspective*. The second has been called the problem of *other minds*, or the *third-person perspective*.

3. Another important issue is *semantics*, which deals with the problem of the *meaning of mental concepts* and the *methodology* that tries to determine the *best methods for the study of consciousness phenomena.*[5] In the area of methodology, the importance of the cooperation between philosophy and natural science is illustrated best.

2. The analysis of theories of consciousness

The analysis of philosophy has shown that positions represented for example by theses of *substance dualism*, *semantic physicalism*, *functionalism, and identity theory* have *many weaknesses*, so *they can impossibly demonstrate what they intend to.* Other theories, like those of Frank Jackson, Thomas Nagel, Joseph Levine have led to *philosophical progress* by contributing to a *better understanding* of the various and complex aspects of the problem of consciousness.[6] They pointed out that it is conceivably *not possible to reduce phenomenal states to objective physical states* and that *they cannot be identical* with brain states. According to these authors, *there is no conceptual or analytical link between the concepts of*

[2] Thomas Metzinger, *Bewusstsein. Beiträge aus der Gegenwartsphilosophie* (Paderborn: mentis Verlag, 2005), 17-21.

[3] Ansgar Beckermann, *Analytische Einführung in die Philosophie des Geistes* (Berlin: Walter de Gruyter, 2001), 1-2.

[4] Beckermann, *Analytische Einführung*, 2.

[5] Beckermann, *Analytische Einführung*, 2-3.

[6] Beckermann, *Analytische Einführung*, 429-430.

consciousness and the physical, non-mental concepts, by which consciousness can be explained or reduced. However, the remaining question, regarding the *nature of this link*, remains open.

3. Phenomenal Consciousness and *„Qualia"*

Many philosophers improperly used the term '*consciousness'* to mean the *inner spiritual world*, which is similar to the *physical inner world*, even if they differ fundamentally.[7] Consciousness mainly implies *subjective experience*, so it has *phenomenal characteristics* or *qualities of experience*, also called *'qualia'* in philosophy.

In *the subjective nature of mental states* the main concern is, according to Nagel's famous essay,[8] *"what it feels like"* (e.g. "*what it is like to be a bat,"* which has different sensory organs in comparison to other species). Another well-known example is that of Jackson[9] about Mary, who is a specialist in the research field of perception, but does not have the ability of color perception. When Mary leaves her black-white-gray prison and for the first time sees a ripe tomato, she has acquired something new. These essays are far from being able to define the complex field of phenomenal consciousness, and generally *a non-circular definition of consciousness cannot be avoided*. One can only explain this aspect by using synonymous terms and referring to examples.[10] However, they offer explanations that serve to avoid confusion with other applications of the concept of consciousness.[11]

4. The necessity for the integration of two types of data for the study of consciousness

Metzinger[12] argues that ultimately a *good theory of consciousness* has to be accepted as a *theory of our own inner experience*. It needs to account for the *subtlety and* phenomenological *richness of experience* and to really *take seriously the internal perspective* of the experiencing subject. Moreover, it has to explain to us *how the*

[7] Beckermann, *Analytische Einführung*, 13.

[8] Thomas Nagel, "What is it like to be a bat?" *Philosophical Review* 83 (1974): 435-450.

[9] Frank Jackson, "Epiphenomenal qualia," *Philosophical Quarterly* 32 (1982): 127-136.

[10] Beckermann, *Analytische Einführung*, 384.

[11] Ned Block, "Eine Verwirrung über eine Funktion des Bewusstseins", in *Bewustsein. Beiträge aus der Gegenwartsphilosophie*, Hrsg. Thomas Metzinger (Paderborn: Mentis Verlag, 2005), 523-581.

[12] Metzinger, *Bewusstsein*, 18.

first-person perspective is related to the third-person perspective of the externally operating science.

Chalmers[13] approached these issues and demonstrated that the *research progress in the field of cognitive psychology and neuroscience requires consideration of both the first-person perspective* and a *third-person perspective.*

Third-person *data* present *neutral phenomena* by reflecting *behavioral data* and data *on brain processes*. They provide traditional material for *cognitive psychology and neuroscience* with the phenomenal aspect remaining unexplained.

First-person *data* are *subjective*, since they are concerned with *data about emotional experiences*. They provide a second perspective for the science of consciousness and allow *access to the phenomenal experience*. However they make no statement about cognitive and neural mechanisms.

Consequently, *first-person data cannot be reduced to third*-person *data and vice versa*. This means that third-person data alone provide an incomplete data catalog, since the phenomenal aspect remains unexplained. Only first-person data are also incomplete, since they make no statement about cognitive and neural mechanisms. This is the *explanatory gap*, which was discovered in the current state of the art regarding knowledge in the area of consciousness. One can therefore say that *the association* between *objective functions and a certain kind of subjective experience requires* the integration of *both kinds of data*. This would be the main objective of a satisfactory study of consciousness, which would allow building an explanatory bridge in a scientific context.[14] Both data types require explanation and interpretation.

5. First-person data: the "difficult" problem of consciousness

In philosophy, the problem connected with *the explanation of the third-person data of consciousness* is also known as *the 'simple' problem of consciousness*, since *clear methods* of implementation for collecting such data are directly available among standard procedures of the *cognitive psychology and neurosciences*. In this way, the processes are discovered and specified in terms of *computational and neural mechanisms*. Chalmers[15] indicates that third-person data explain how the system is objectively functioning. A *reductive explanation model* (e.g. higher-level phenomena can be explained by low-level processes i.e. molecular biology) can

[13] David J. Chalmers, "How can we construct a science of consciousness?" in *The Cognitive Neurosciences III*, ed. Michael S. Gazzaniga, *The Cognitive Neurosciences III* (Cambridge: MIT Press, 2004), 1111-19.

[14] Chalmers, "How can we construct."

[15] Chalmers, "How can we construct."

only be useful for clarifying the *objective function of the cognitive system* in the form of neurophysiology.

This model is not appropriate for *first-person data* because such data deal with *subjective experience*. A complete report on the objective functions of consciousness cannot possibly answer questions on the association of these functions with a certain kind of subjective experience. This problem can *withstand these methods*. The *problem of explaining the first-person data of consciousness* is sometimes called *the 'difficult' problem of consciousness*. According to Chalmers, important questions, even after completing the picture of the objective functions of the brain and behavior, remain unanswered. In general, can one tackle this problem at all with the tools of neuroscience? Why are these functions associated with conscious experience? And why are they connected with a certain kind of experience?

The *obstacles in the collection of first-person* data are significant, e.g. the privacy of this kind of data; the *lack of inter-subjective perception* (there is no measure of consciousness); the *less advanced development of methods* for the investigation of first-person data in comparison with those for collecting third-person data, particularly with concern to more subtle phenomena; *lacking formal criteria and theory development* for collecting first-person data.

6. Integration modalities for both kinds of data

Chalmers[16] made *methodological suggestions* for further research in this area. He proposed the following: *correlate detailed first-person features with third-person features*, *systematize the connection* with principles of increasing generality, and use preferably *simple, basic, and universal principles* that underlie and explain the higher-level connections.

He suggests that one can *solve these problems in a roundabout way*: e.g. *comparing conscious-unconsciou*s, finding *behavioral and neural correlates of subjective experiences*, monitoring *subjective verbal reports*, applying observational methods. The main types of first-person data are based on *visual perception* (e.g. the perception of color and depth), *other senses* (e.g. hearing and tactile sense), *bodily sensations* (e.g. pain and hunger), imagination (e.g. memory of visual images), *emotional experiences* (e.g. happiness and anger), and *thoughts* (e.g. deliberations and decisions). Nevertheless, the connection between these two types of data may require a theory based on *principles of structural coherence, organizational invariance, and double perspective*.

[16] Chalmers, "How can we construct."

As for *verbal reports*, philosophers like Chalmers or Metzinger refuse to recognize them as first-person data. Metzinger has gone further and claimed that there are no first-person data. This is a very challenging position, which precludes the hope of a scientific approach to this data. Data are considered *objective* if they *proportionally correspond to the measured aspect of reality*. As a parenthesis, *all data are actually subjective*, but if the subjective estimation is confirmed in a variety of situations, then we can consider the measure as objective. The *example of temperature measurement* has been frequently used to illustrate the subjectivity of data: we measure the temperature not directly, but we assume that the highness of the mercury column is proportional with the temperature.

In the case of *verbal reports*, there are procedures in which one looks at many different subjective experiences for a *large number of persons and conditions* and *verifies this data by objective instrumental measures*, an option that allows *access to the first-person data*. If the received *verbal reports are proportional to the subjective experience*, they allow *access to valid first-person data*. Here I find the suggestion of Hobson[17] constructive: he proposes a compromise by which verbal reports can be at least considered *"third person half-some-one'."*

7. Implicit and explicit memory

So far I have presented views and suggestions coming from the field of the philosophy of mind. Using the *example of the memory research*, I will now show that *first-person data have been often less considered in the cognitive psychology and neuroscience*.

In the philosophy of mind, *memory* can be commonly referred to as an *information supplier* from which stored information can be retrieved. In psychology, memory is defined as the *brain's ability to receive, retain, organize, and retrieve information*. The memory content is set out in the *synaptic efficiency of neural networks* and the dominant metaphor for memory retrieval is the *association*. In this way, words, phrases, and also emotions are seen as part of a large network, with its adjacent areas being semantically related to each other.

Regarding the relationship between memory organization, brain structures, and memory capacity, the fundamental idea in psychological literature is that *memory is not a single unit*. Apparently memory consists of *several separate and partially independent components that rely on different brain systems*. The most common types of distinctions are made between *long- and short-term memory* and between *implicit and explicit memory*. The explicit/*declarative* memory is the

[17] Allan Hobson, "Finally Some One: Reflections on Thomas Metzinger's 'Being No One,'" *Psyche* 11, 5 (2005): 1-20.

capacity for *memory of facts/general-knowledge information (semantic memory)* and *events/situations (episodic memory)*, which can be *deliberately retrieved* and *verbally reported within a chronological context.*[18] The processes of declarative memory appear to be based on the activity of the *hippocampus* and adjacent cortical structures, and of the frontal lobe.[19]

The *implicit memory* is conceived as a *heterogeneous collection of unconscious learning skills* (i.e. *non-declarative/procedural memory*) that are expressed through *performance* and for which *access to any conscious memory* content is mostly not available. It refers to *habits, skills, emotional reactions, reflexes and conditioned responses*. These processes are linked to different specific areas of the nervous system, e.g. *amygdala.*[20] Implicit memory contents are *activated by cues* and characterized by sensory, emotional, and physiological perception accompanied by feelings of '*here and now,'* and *could not be verbally reported in a coherent form or logically explained*.

The division between implicit and explicit memory was initially based on the evidence that such processes are *experimentally separable*. Studies showed that performance improvements in fulfilling tasks and the ability to learn are possible without conscious recollection of the learning episodes in amnestic patients. Thus, the processes and the relevant areas of implicit memory seem too heterogeneous to be included in a unitary memory system.[21] Moreover, the subdivision of implicit memory seems to correspond rather to the types of tasks than to the criteria of consciousness.[22] Brain studies have indicated that these memories are processed in *different brain areas*.[23] There is also evidence that *glucocorticoids* have an important role in the regulation of imprinting, consolidation and retrieval of *declarative memory*.[24] The *amygdala* modulates the strength of both declarative and procedural memory processes.

[18] Martin A. Conway and Christopher W. Pleydell, "The construction of autobiographical memories in the self-memory system", *Psychological Review* 107, 2 (2000): 261-288.

[19] Larry R. Squire, "Declarative and nondeclarative memory: multiple brain systems supporting learning and memory", in *Memory systems*, eds. Daniel L. Schacter and Endel Tulving (Cambridge: MIT Press, 1994), 203-232.

[20] Squire, "Declarative and nondeclarative memory," 215-224.

[21] Daniel B. Willingham, Laura Preuss, "The death of implicit memory," *Psyche* 2, 15 (1995), http://psyche.cs.monash.edu.au/v2/psyche-2-15-willingham.html.

[22] Willingham, Preuss, "The death of implicit memory."

[23] Squire, "Declarative and nondeclarative memory," 215-224.

[24] See C. Kirschbaum, O. T. Wolf, M. May, W. Wippich, D.H. Hellhammer, "Stress- and treatment-induced elevations of cortisol levels associated with impaired declarative memory in healthy adults," *Life Sciences* 58 (1996): 1475-1483; Dominique J.-F. de Quervain, Benno

8. The neglect of first-person data in memory research

A *key criterion* for distinguishing between explicit and implicit cognitive functions is the *presence / absence of conscious knowledge*. Implicit memory is demonstrated when performance by fulfilling a task is facilitated by the absence of conscious recollection. Explicit memory is demonstrated when performance requires conscious recollection of previous events/knowledge.

The *methods* used for testing implicit processes are *different* from those that assess explicit functions. For example, *declarative memory* is directly tested by asking participants to *consciously recall something*. On the other hand, *implicit memory* is usually studied by *evaluating performances* depending on indirect recall and expressing behavioral changes.

Gardiner[25] investigated the direction research on implicit memory leads into. A picture of the restrictions in this area is presented by *two main methods* preferred by many researchers: the *criterion of intentional retrieval*[26] and the *procedure of process dissociation*.[27] In the intentionality criterion of retrieval are reached conclusions about the nature of implicit versus explicit memory by experimental designs, in which the same stimuli for implicit and explicit tests are given to participants and only test instructions vary. The procedure of process dissociation investigates the cognitive control of the recall for the completion of word stems during two experimental conditions. In the first condition, participants are requested to complete the stems from a previously studied list of words. In the second condition, participants are asked to complete the word stems that were not on the list. Gardiner criticizes this approach, since it falls under the category of *third-person explanations of consciousness*. He proposes the use *experiential procedures* aiming at correlating this data with first-person data.

Roozendaal, James L. McGaugh, "Stress and glucocorticoids impair retrieval of long-term spatial memory," *Nature* 394 (1998): 787-790; J. W. Newcomer, G. Selke, A. K. Melson, T. Hershey, S. Craft, K. Richards, A. L. Alderson, "Decreased memory performance in healthy humans induced by stress-level cortisol treatment," *Archives of General Psychiatry* 56 (1999): 527-533; Werner Plihal, Jan Born, "Memory consolidation in human sleep depends on inhibition of glucocorticoid release," *Neuroreport* 10 (1999): 2741-7; Benno Roozendaal, "Glucocorticoids and the regulation of memory consolidation," *Psychoneuroendocrinology*, 25 (2000): 213-238.

[25] John M. Gardiner, "Functional aspects of recollective experience," *Memory and Cognition* 16 (1988): 309-313.

[26] Daniel L. Schacter, Jeffrey Bowers, Jill Booker, "Intention, awareness, and implicit memory: The retrieval intentionality criterion," in *Implicit memory: Theoretical issues*, eds. Stephan Lewandowsky, John C. Dunn, and Kim Kirsner (Hillsdale: Erlbaum, 1989), 47-65.

[27] Larry L. Jacoby, "A process dissociation framework: Separating automatic and intentional uses of memory," *Journal of Memory and Language* 30 (1991): 513-541.

Kihlstrom adds more criticism by showing that *such experiments do not really investigate what they intend to investigate*. Although the tasks involve complex rules that are unknown and unpredictable for the participants, *volunteers gain explicit knowledge* during the experiment, which helps them fulfill the tasks and can explain their performance. Kihlstrom et al.[28] propose a different approach in the investigation of unconscious memory processes.

9. Interactions between the implicit and explicit memory processes

However, it seems *oversimplified* to divide memory into two precise mental entities. The *interaction between implicit and explicit processes* and mechanisms has been *demonstrated experimentally*. For example, in the case of *perceptual-motor skill learning*[29]: This is a test in which 4 lines continuously appear on the lower part of a screen to indicate sites where asterisks may appear. At each appearance of an asterisk, participants must press the corresponding button (A, B, C or D). Participants were naive with respect to the existence of sequence of appearance sites, which is repeated again and again. Results showed improved response times for both amnestic patients and in healthy controls. However, healthy participants acquired a certain degree of declarative knowledge during the tests. Similar findings also came from other tests for implicit memory and *partially overlapping brain activation patterns during implicit and explicit processing have been recently demonstrated*. The *earlier view* that completely different brain areas are responsible for implicit and explicit memory processes became *invalidated*. This shows that *these processes work simultaneously and harmoniously, and it is difficult to distinguish between them.*[30]

Hence the question arises: Is the distinction valid and useful? Are there qualitative / quantitative differences? Related conceptual criticism questioned the categorical division between implicit and explicit cognitive processes, proposing that apparently intact implicit processing in the presence of apparently disturbed explicit processing may solely reflect "*a more degraded modus operandi of the cognitive system as a whole, i.e. with 'explicit' and 'implicit' processes in fact lying*

[28] John F. Kihlstrom, Terrence M. Barnhardt, Douglas J. Tataryn, "The psychological unconscious. Found, lost, and regained," *American Psychologist* 47 (1992): 788-91.

[29] Paul J. Reber, Larry R. Squire, "Parallel brain systems for learning with and without awareness," *Learning & Memory* 1 (1994): 217-229.

[30] Deborah Faulkner, Jonathan K. Foster, "The decoupling of 'explicit' and 'implicit' processing in neuropsychological disorders: insights into the neural basis of consciousness?" *Psyche* 8, 2 (2002), http://psyche.cs.monash.edu.au/v8/psyche-8-02-faulkner.html.

on a functional continuum."[31] Consequently, the apparent *preservation of qualitatively separate implicit / non-conscious brain units* in amnestic patients *may embody the preservation of simpler and less resource-consuming processing patterns*. In this way, the question about the nature of conscious knowledge remains open.

10. The improper use of the concept of 'consciousness' in memory research

Another problem is that phenomenal consciousness (*P-consciousness*) is often *confounded with other types of consciousness, usually with access-consciousness.*[32] So speaks, for example, Baars[33] about the nature of experience (P-consciousness), but his theory is a *"global workspace" model of access-consciousness*. Also, Jacoby et al.[34] argued that studying the processes of access-consciousness may say a lot about P-consciousness.

Baars[35] argues that while Nagel's criterion is a too demanding criterion for an empirical science of consciousness, behavioral denial of the phenomenal aspects of consciousness research is too restrictive and this endless debate is fruitless. He proposes a *compromise for consciousness research by specifying comparable pairs of psychological phenomena*, which differ in only one point. One part is aware and the other is not (e.g. conscious / unconscious memory). He calls this method a *"method of contrastive analysis."*

11. Examples of knowledge advance on memory processes coming from the field of clinical psychology

The field of *clinical psychology necessarily relies* much more than other psychology domains *on subjective, first-person data*. The research of emotion is devoted to *studying the behavior and physiology of human emotion* by means of which we are able to observe and understand humans. For this kind of research, it is important to *scan, consciously integrate and probe information from three levels: subjective* (verbal expression, prosody), *behavioral* (motor, facial expression, etc.), *physiological*

[31] Faulkner, Foster, "The decoupling of 'explicit' and 'implicit' processing in neuropsychological disorders," 4.

[32] Block, "Eine Verwirrung."

[33] Bernard J. Baars, "A Thoroughly Empirical Approach to Consciousness," *Psyche* 1, 6 (1994), http://psyche.cs.monash.edu.au/v2/psyche-1-6-baars.html.

[34] Larry L. Jacoby, D. Stephen Lindsay, Jeffrey P. Toth, "Unconscious influences revealed. Attention, awareness, and control", *American Psychologist*, 47 (1992): 802-809.

[35] Baars, "A Thoroughly Empirical Approach."

(trembling, sweating, and crying). Lang[36] emphasized that *assessments that leave out one or more of these three modes of emotional expression can be highly misleading*. Contrasting research in other areas, particularly *psychotraumatology research*, has employed *multimodal assessment strategies*: first-person data (e.g. data coming from the observation of behavior, self-report measures, and clinical interviews) and third-person data (neuropsychological tests, learning experiments, psychophysiological and brain investigations). This approach of clinical psychology combining both types of data led to significant advances of knowledge in the area of consciousness. Here I will exemplify such important *advances in the area of implicit and explicit memory processes* with important implications for research, clinical practice, and psychotherapy.

Memory processes are of great relevance in the area of stress-related disorders (e.g. posttraumatic stress disorder/PTSD, dissociative disorders, and borderline personality disorder). The *response to stress* appears to be mediated by specific *neurochemical and neuroanatomical dysfunctions*[37]: When stress increases, both the *hippocampus and the amygdala increase their activity* and *stress hormones* (e.g. adrenalin, noradrenalin, and glucocorticoids) are being released into the circulatory system. From a certain point onward, when the level of stress is very high, the *hippocampus becomes less functional* and the *amygdala reaches a plateau level*. It has been showed that *elevated doses of glucocorticoids* impair hippocampal activity and *have damaging effects on the hippocampus on the long-term*: atrophy and loss of pyramidal neurons, reduction of the ramifications of the hippocampal dendrites and that adrenalin and noradrenalin in high concentrations increase the activity of the amygdala (particularly in case of chronic/prolonged trauma[38]). *Hippocampal dysfunctions* are thought to impair the *encoding of explicit information* and, subsequently, the *access to the elements of the trauma-related explicit memory*. It is assumed that this *dysfunctional encoding of distressing/traumatic events* is the way in which the *posttraumatic symptoms are generated*.

The *"fear network" model* of trauma-related memories originating in the work of Lang[39] states that a *sensory-perceptual representation* including elements of

[36] Peter J. Lang, "What are the data of emotion?" in *Cognitive Perspectives on Emotion and Motivation*, eds. V. Hamilton, G. H. Bower, and N. Frijda (Boston: Martinus Nijhoff, 1988).

[37] Bessel A. van der Kolk, "The psychobiology of posttraumatic stress disorder," *Journal of Clinical Psychology* 58 (1997): 16-24.

[38] James L. McGaugh, "Significance and remembrance: The role of neuromodulatory systems," *Psychological Science*, 1 (1990): 15-25.

[39] Peter J. Lang, "A bio-informational theory of emotional imagery," *Psychophysiology* 16 (1979): 495-512.

implicit memory (i.e. peritraumatic strong bodily sensations, intensive emotions, thoughts, and behavioral reactions) is formed, but these elements are *not well integrated in the autobiographical/explicit memory* at the same time. This fear network is *highly consistent, very large, and long-lasting*, has particularly *strong links* and can be *easily activated*. By contrast, there are fewer activation pathways going from the implicit memory system to the elements of explicit memory (i.e. knowledge of general, specific events and lifetime periods). Consequently, the *autobiographical representation* (explicit memory) is highly fragmented, inconsistent, includes partial amnesia, contradictory information.

For the psychophysiological and neurobiological research (third-person data), paradigms including first-person data have been employed: the paradigm of *emotional imagery*, the paradigm of *exposure to trauma-related stimuli*, *subjective ratings* of emotional experiences during event. In the paradigm of emotional imagery (*script-driven imagery*), during the reading of a personalized report of a traumatic event (*script*), participants are asked to vividly imagine this situation including actions, persons and emotions present during the real situation (imaginative procedure). The paradigm of exposure to trauma-related stimuli employed various material (auditive, visual, and combinations of auditive and visual stimuli).

Several results emphasize the importance of this multimodal approach. Findings of experiments using imagery and exposure to trauma-related stimuli indicated a higher physiological reactivity (e.g. heart rate, skin conductance, blood pressure, muscular activity, activation of certain brain areas, amplitude of the blink reflex) of PTSD patients as compared to traumatized persons without PTSD and to controls. Other studies proved that physiological reactivity to stimuli related to trauma may predict the development and persistence of PTSD.[40] The physiological reactivity to stimuli related to trauma may allow for the evaluation of treatment efficiency.[41] Hippocampal Magnetic Resonance Imaging studies indicated lower hippocampal volumes in PTSD patients.[42] The meta-analysis of clinical and

[40] See E.B. Blanchard, E.J. Hickling, A.E.Taylor, W.R. Loos, C.A. Forneris, J. Jaccard, "Who develops PTSD from motor vehicle accidents?" *Behaviour Research and Therapy* 34 (1996): 1-10; A.Y. Shalev, T. Sahar, S. Freedman, T. Peri, N. Glick, D. Brandes, S.P. Orr, R.K. Pitman, "A prospective study of heart rate response following trauma and the subsequent development of posttraumatic stress disorder," *Archives of General Psychiatry* 55 (1998): 553-559.

[41] Scott P. Orr, Walton T. Roth, "Psychophysiological assessment: clinical applications for PTSD," *Journal of Affective Disorders* 61 (2000): 225-240.

[42] J.D. Bremner, P. Randall, E. Vermetten, L. Staib, R.A. Bronen, C. Mazure, S. Capelli, G. McCarthy, R.B. Innis, D.S. Charney, "Magnetic resonance imaging-based measurement of hippocampal volume in posttraumatic stress disorder related to childhood physical and sexual abuse–a preliminary report," *Biological Psychiatry* 41 (1997): 23-32.

neurobiological indicators for PTSD[43] demonstrated the existence of mainly *two brain activation patterns* corresponding to *two pathological PTSD subtypes*: (1) the dissociative subtype of PTSD characterized by emotional inhibition and by an inhibition of the limbic system through the activation of the medial prefrontal cortex; (2) the PTSD subtype characterized by emotional activation (predominant re-experiencing/hyperarousal symptoms), mediated by the failure of prefrontal inhibition of the same limbic regions.

These findings *inspired the development of effective treatment strategies*, suggesting that the *trauma-related fear network (implicit memory) should be activated* and *durably modified* by *adding new elements* that are incompatible with the original pathological memory representation. Consequently, most successful therapies operate an *integration of explicit/declarative memories within a coherent autobiographic report* (through repeated exposure to traumatic memories) and *correct dysfunctional old trauma-related beliefs by means of cognitive restructuring*. In this way, the pathological fear response is inhibited. Findings demonstrating that these *old fear responses can be reactivated in certain situations* even after successful therapy[44] *document the real neurobiological substrate of treatment effects*. This is not about a new memory restructuring (deletion of the old neural synapses of the fear network), but about a *memory restructuring though the learning of new elements*. On a neuroanatomical level, the extinction of fear response is mediated by the inhibitory influences of the medial prefrontal cortex on the amygdala, which has been confirmed by the research of brain activation patterns. Accordingly, one could say the following: (a) trauma-focused therapy leads to the inclusion of neutral, declarative contents into the memory system to form a new trauma-associated memory representation; (b) the new memory contents are parallel to the original ones and inhibit, depending on the context, the activation of the still intact old fear structure during the confrontation with trauma-associated stimuli/situations.

Multimodal assessments proved their *utility* in several ways. This approach demonstrated its *high efficacy for knowledge advance* as well as *clinical diagnostic and therapeutic value* by furnishing an *in-depth understanding of neurobiological substrates and mechanisms of psychological disturbances and their treatment*.

[43] R.A. Lanius, E. Vermetten, R.J. Loewenstein, B. Brand, C. Schmahl, J.D. Bremner, D. Spiegel, "Emotion modulation in PTSD: Clinical and neurobiological evidence for a dissociative subtype," *American Journal of Psychiatry* 67 (2010): 640-647.

[44] See S. Rachman, M. Whittal, "The effect of an aversive event on the return of fear", *Behaviour Research and Therapy* 27 (1989): 513-520; S. Rachman, M. Whittal, "Fast, slow and sudden reductions in fear," *Behaviour Research and Therapy* 27 (1989): 613-620.

12. Explanatory gaps in memory research

Research in the field of implicit and explicit memory has led to *insights about the objective functions* in terms of *cognitive processes and responsible brain areas*. Due to the *preference for third-person methods* by cognitive researchers and neuroscientists, most frequently *accounts of subjective experience (first-person data) have been left aside*. Although scientists in the field of memory research aim at investigating phenomenal consciousness, they have used *no solid principles for the collection of first-person data*. Most frequently, the aspects of phenomenal consciousness have been confounded with other types of consciousness. Previous findings in this area say little about the phenomenal aspects of consciousness accompanying memory processes. They provide even less evidence on how these functions are associated with subjective experiences. Some areas of *clinical psychology research have overcome some of these deficits* by using multimodal data collection that led to *significant knowledge advances in the area of explicit and implicit memor*y. However, since *clinical research significantly relies on fundamental research* by approving and integrating concepts and findings coming from the cognitive psychology and neurosciences, many questions about the validity of implicit-explicit distinction and about the nature of conscious knowledge still remain unanswered.

Despite the advances in cognitive psychology and neurosciences, the problem remains unsolved. One reason for the strong increase in the current interest in the exploration of consciousness lies in the development of new technologies for brain research. This has led to a widespread optimism among neuroscientists in terms of access to a theory of consciousness. Many philosophers who have been following and have praised this progress of the neural correlates of consciousness, however, have realized that the explanatory gap remains unsolved. The question is what kind of theory and what types of methods are needed for building a bridge between the neural correlates and phenomenal elements of experience?

13. Proposals for a memory research that emanates from the unitary existence of phenomenal consciousness

Following *constructive philosophical recommendations* and *positive examples coming from areas of clinical psychology research*, I suggest that cognitive and neurophysiological fundamental research should make more intensively and systematically use of first-person data, which should be combined with third-person data. I propose the following:

- It should *distinguish more strictly between working concepts*, develop *theories about the function of consciousness* and related, testable research hypotheses. This should be achieved through *more rigorous* *classification criteria* and the operalization *of* memory and learning *concepts*.

- It should develop more *sensitive methods* for the investigation of first-person data (e.g. sensitive scales and formal and content data analysis of verbal self-reports and benchmarks for observation that should be collected additionally to third-person data).
- It should develop *valid procedures for the correlations* between the two types of data (e.g. isomorphism between aspects of phenomenological consciousness and changes in brain activity that occur simultaneously). An *interpretation of the associations as neuronal correlates* of certain aspects of consciousness should be *subsequently tested* on persons with neurological and psychological disturbances.

KNOWING FUTURE CONTINGENTS

Ezio DI NUCCI

ABSTRACT: This paper argues that we know the future by applying a recent solution of the problem of future contingents to knowledge attributions about the future. MacFarlane has put forward a version of assessment-context relativism that enables us to assign a truth value 'true' (or 'false') to future contingents such as "There Will Be A Sea Battle Tomorrow." Here I argue that the same solution can be applied to knowledge attributions about the future by dismissing three disanalogies between the case of future contingents and the case of knowledge attributions about the future. Therefore none of the traditional conditions for knowledge can be utilized to deny that we know the future, as I argue in the last section.

KEYWORDS: future contingents, knowledge attributions about the future, John MacFarlane, assessment-context relativism

We know the future: this paper is going to demonstrate it. Whether or not the thesis of determinism is true, we know the future. Whether or not the future is genuinely open, we know the future. By applying MacFarlane's[1] recent solution of the problem of future contingents to knowledge attributions, this paper shows that we know the future.

"The Man Who Will Get The Job Has Ten Coins In His Pocket": you won't find an epistemologist who is not familiar with this proposition. It was famously deployed by Gettier[2] to refute the traditional tripartite analysis of knowledge as true justified belief. Gettier's proposition is peculiar in one respect: it isn't easy to see whether and how one could evaluate it as true or false because it contains a future fact. Whether or not (and how) a truth-value can be assigned to so-called *future contingents* has boggled the minds of philosophers since Aristotle's "There Will Be A Sea Battle Tomorrow."

Future contingents should be of particular interest to epistemologists too: Gettier's counterexample consisted in putting forward a proposition, "The Man Who Will Get The Job Has Ten Coins In His Pocket," which is TRUE, BELIEVED by Smith, and JUSTIFIED. Still, intuitively Smith does not know it – and therefore the traditional three conditions on knowledge are not sufficient. This is because

[1] John MacFarlane, "Future Contingents and Relative Truth," *The Philosophical Quarterly* 53 (2003): 321-36.

[2] Edmund Gettier, "Is justified true belief knowledge?" *Analysis* 23 (1963): 121-23.

Smith has been told by the company's CEO that Jones will get the job, and Smith has also seen that Jones has ten coins in his pocket. Therefore Smith is justified in believing that the man who will get the job has ten coins in his pocket. In fact, it is Smith himself who will get the job – and Smith has ten coins in his pocket too!

But now we can easily see that Gettier's proposition constitutes a counterexample to the analysis of knowledge as true justified belief only if Gettier's proposition is indeed true. But the proposition appears to be a future contingent: therefore it can constitute a counterexample only if we resolve the issue of future contingents in such a way that we can assign a truth-value to future contingents; and that truth-value will have to be TRUE (or FALSE) rather than some third truth-value (i.e. Lukasiewicz's *indeterminate*).

But Gettier's counterexample isn't too much of a problem, given that there are plenty of Gettier-type counterexamples which do not involve future contingents. There is a much more general epistemological question posed by the issue of future contingents: the fact that we claim to know future contingents all the time, and that we often act upon our knowledge and other people's knowledge of future contingents.[3] And when we do so, we behave perfectly rationally. What needs vindicating then is knowledge attributions whose content is a future contingent[4]: I Know That The President Will Be In The Fourth Vehicle Of The Motorcade, you can imagine a conspirator say. Not only does the conspirator's speech not sound weird or inappropriate, but whether or not the conspirator does in fact know makes quite a difference![5]

Here's another example from Goldman:

> Let us grant that I can know facts about the future... T intends to go downtown on Monday. On Sunday, T tells S of his intention. Hearing T say he will go downtown, S infers that T really does intend to go downtown. And from this S concludes that T will go downtown on Monday. Now suppose that T fulfils his intention by going downtown on Monday. Can S be said to know that he would go

[3] More on this in Section III.

[4] Smith would have known that the man who will get the job has ten coins in his pocket if Jones rather than Smith himself had gotten the job – this suggestion is implicit in Gettier's counterexample.

[5] A clarificatory point: when I talk of vindicating our knowledge of future contingents, one should bear in mind the distinction between knowledge attributions being warranted and knowledge attribution statements being true. The former, differently from the latter, doesn't depend on the truth condition on knowledge being satisfied. But the former, differently from the latter, does not imply that the agent to which knowledge is being attributed does in fact know.

> downtown? If we ever can be said to have knowledge of the future, this is a reasonable candidate for it.[6]

Recently MacFarlane[7] has put forward a proposed solution to the problem of future contingents which could be deployed to vindicate our knowledge of future contingents. I turn to this attempt in the next section.

I

MacFarlane puts forward a version of truth-relativism which promises to be able to assign the truth-value true (or false) to future contingents without sacrificing what MacFarlane calls the *indeterminacy intuition*: the idea that the future is genuinely open.

> On the other hand, there is a strong temptation to say that the assertion does have a definite truth-value, albeit one that must remain unknown until the future 'unfolds'. After all, once the sea battle has happened (or not), it seems quite strange to deny that the assertion was true (or false). I shall call the thought that the assertion does have a definite truth-value 'the determinacy intuition.'[8]

MacFarlane's account aims to accommodate both the 'indeterminacy intuition' and the 'determinacy intuition.' On his view, truth is relative to its context of assessment. In the case of future contingents such as "There Will Be A Sea Battle Tomorrow," then, the statement will be true when assessed from a future context – say tomorrow in the midst of battle. When assessed today, the statement is neither true nor false. This gives us a way of saying that future contingents are true (or false); they can be true when assessed from a particular context. But if truth is indeed relative to the statement's context of assessment, then future contingents are not special cases: the only sense in which any statement is ever true is that it is true as assessed from a particular context, according to this assessment context-relativism about truth.[9]

So not only does MacFarlane offer a solution to the problem of future contingents; MacFarlane offers the kind of solution we can help ourselves to in order to vindicate our knowledge of future contingents. MacFarlane offers us a way of meeting the truth-condition on knowledge: "The Man Who Will Get The Job Has Ten Coins In His Pocket" is true when assessed from a context such as a time *after* the CEO has offered the job to Smith.

[6] Alvin I. Goldman, "A Causal Theory of Knowing," *The Journal of Philosophy* 64 (1967): 364-65.

[7] MacFarlane, "Future Contingents and Relative Truth."

[8] MacFarlane, "Future Contingents and Relative Truth," 321

[9] I should emphasize that my argument is conditional: I am not going to defend MacFarlane context-relativism about truth; I will just show what we can do with it.

What we end up with, then, is the extension of MacFarlane's relativism to knowledge attributions. As assessed from a later context, knowledge attributions which contain future contingents are also true. Suppose that tomorrow our conspirator targets the fourth vehicle in the motorcade, killing the President, who was indeed travelling in the fourth vehicle. Now we can say that our conspirator knows, today, that the President will be in the fourth vehicle, as assessed from tomorrow night's context of assessment.

Just as "There Will Be A Sea Battle Tomorrow" is true as assessed from a later context (tomorrow, in the midst of battle), in the same way "The Conspirator Knows That The President Will Be In The Fourth Vehicle" is also true as assessed from a later context (tomorrow evening while America is in mourning, say). The 'price to pay' is relativism about knowledge attributions (and indeed MacFarlane has independently argued for relativism about knowledge attributions[10]). But it is only natural to think that, if TRUTH is context-relative, then knowledge attributions will also be context-relative – at least if we think that KNOWLEDGE implies TRUTH.

So even though epistemologists might not be willing to concede relativism about knowledge attributions in order to vindicate our knowledge of future contingents, relativism about knowledge attributions simply follows from MacFarlane's general context-relative account of truth – as long as we are unwilling to give up on the TRUTH condition on KNOWLEDGE.[11]

II

In this section I point to three important disanalogies between the context-relative truth of "There Will Be A Sea Battle Tomorrow" and the context-relative truth of "The Conspirator Knows That The President Will Be In The Fourth Vehicle." These disanalogies must be overcome if we are to successfully vindicate knowledge of future contingents.

First disanalogy

A later context of assessment is the *proper* context of assessment in the Sea Battle case, but it is not the *proper* context of assessment in the Conspirator case.

[10] John MacFarlane, "The Assessment Sensitivity of Knowledge Attributions," in *Oxford Studies in Epistemology* 1, eds. Tamar Szabò Gendler and John Hawthorne (Oxford: Oxford University Press, 2005), 197–233, "Relativism and Knowledge Attributions," in *Routledge Companion to Epistemology*, eds. Sven Bernecker and Duncan Pritchard (London: Routledge, 2010), 50.

[11] There is an independent way in which assessment context-relativism about knowledge attributions follows from assessment context-relativism about truth: knowledge attributions are assessment context-relative simply because all propositions are.

MacFarlane does not talk of proper or appropriate contexts as opposed to inappropriate contexts. But the strength of his proposed solution of allowing for future assessment of future contingents appears to derive from the fact that the more appropriate context of assessment for a statement about tomorrow is indeed tomorrow. So that the statement "There Will Be A Sea Battle Tomorrow" is saying something about tomorrow and should be assessed tomorrow. But this isn't the case for knowledge attributions that contain future contingents: "The Conspirator Knows That The President Will Be In The Fourth Vehicle" does not say something about tomorrow; or, anyway, it does not *only* say something about tomorrow. It says, importantly, something about today, namely that the conspirator knows, *today*, where the President will be tomorrow.

Here I don't intend to look at the wider issue of whether the knowledge attribution statement is a *real* future contingent or not.[12] The important point is that, even if it is, it is importantly different from future contingents such as "There Will Be A Sea Battle Tomorrow," because the knowledge attribution (also) describes today's world.

Two points here: crucially, the statement "The Conspirator Knows That The President Will Be In The Fourth Vehicle" says something about today *and* something about tomorrow. So it won't do to only assess it from a present context. That would mean dismissing a crucial aspect of the statement: that it *also* says something about tomorrow. And we will see in the discussion of the third disanalogy that assessing it from a later context does not mean sacrificing what the statement says about today.

Secondly, talking of proper contexts of assessment and improper contexts of assessment (or, for that matter, of more proper contexts than others) betrays the spirit of relativism; we might be unwilling to accept a relativistic proposal in principle; but if we are willing to consider it, then we cannot also take an independent standpoint from which we evaluate the different contexts from the outside.

[12] What stand we take in that wider issue will also determine whether we think that Gettier's "The Man Who Will Get The Job Has Ten Coins In His Pocket" is a *real* future contingent or not: Gettier's statement, one could argue, contains a future contingent (S will get the job), but it is not a statement about the future (a time after the CEO has made the job-offer), because by then S could have taken the ten coins out of his pocket. So there are statements about the present which contain future contingents: knowledge attributions are one example; composite statements such as "The Man Who Will Get The Job Has Ten Coins In His Pocket" are another example. But whether we should also label these kinds of statements *future contingents* isn't crucial to my argument.

Second disanalogy

In the Sea Battle case, the statement is true as assessed from a future context and neither true nor false as assessed from a present context. While in the Conspirator case, the statement is true as assessed from a future context and *false* as assessed from a present context: because knowledge requires truth, the Conspirator knows only if the statement in question is indeed true; but since the statement in question is neither true nor false, then the Conspirator does not know.

This is important because if, as assessed now, the statement "The Conspirator Knows That The President Will Be In The Fourth Vehicle" is false, then the Conspirator does not know, now, where the President will be; and it is now that whether or not she knows will make a difference to her plans. Therefore we haven't actually vindicated our knowledge of future contingents.

But within a relativistic picture it is perfectly fine that a statement is false as assessed from one context and true as assessed from a different context. Also, that the statement is false as assessed from a present context does not mean that the Conspirator does not know *now*. The Conspirator does know now, as assessed from a later context. And the Conspirator does not know now, as assessed from a present context. So there still is a way of vindicating the fact that the Conspirator does know now.

Third disanalogy

If we want to say that the context of KNOWLEDGE corresponds to the context of TRUTH, so that the Conspirator knows as assessed from a later context because, as assessed from *that* context, "The President Will Be In The Fourth Vehicle" is true, then the context of JUSTIFICATION (and the context of BELIEF) must also correspond to the context of TRUTH. But it is not obvious that this will be the case: the Conspirator might be justified in her belief as assessed now and not justified in her belief as assessed from a later context, even if the President does turn out to be in the fourth vehicle.

Suppose, for example, that the source upon which the Conspirator had based her judgement later tips the Conspirator that the President will in fact be in the fifth vehicle. Then the Conspirator would no longer be justified in believing that the President will be in the fourth vehicle, we might suppose, because that source was her only evidence. Still, the Conspirator knows now even if she later changes her mind. But now we can't show that she does know either as assessed from now (TRUTH condition on KNOWLEDGE is not met) or as assessed from tomorrow (JUSTIFICATION and BELIEF conditions are not met).

But while it is true that a present context of assessment is missing the TRUTH condition, it isn't true that a future context of assessment is missing the JUSTIFICATION and BELIEF conditions. Suppose that on Wednesday the President travels in the fourth vehicle. Suppose that at 5pm on Tuesday the Conspirator, having been tipped by an extremely reliable inside source, believes that the President will be in the fourth vehicle, and justifiably so. At 5.01pm, the source tips the Conspirator that, actually, the President will be in the fifth vehicle. So from 5.01pm on Tuesday the Conspirator believes that the President will be in the fifth vehicle, and justifiably so.

As assessed from a present context, we are missing the TRUTH condition, so that we cannot vindicate the statement "The Conspirator knows that the President will be in the fourth vehicle"; but as assessed from a later context (tomorrow after the President has indeed travelled in the fourth vehicle), we are *not* missing the JUSTIFICATION AND BELIEF conditions just because the Conspirator later changes his mind. If we are evaluating the statement that, up until 5pm on Tuesday, the Conspirator knows that the President will be in the fourth vehicle, then we have the TRUTH condition (because we are assessing from a later context); and we have the JUSTIFICATION and BELIEF conditions, because even from a later context of assessment the Conspirator was indeed justified in believing that the President will travel in the fourth vehicle – up until 5pm on Tuesday anyway.

Even though from Wednesday's context of assessment, it is still Tuesday up to 5pm that we are evaluating; so it does not matter that after 5pm on Tuesday the Conspirator is no longer justified.

We have now dismissed three attempts to show that MacFarlane's strategy cannot be applied to "The Conspirator knows that the President will travel in the fourth vehicle." So if MacFarlane's strategy works for standard future contingents, then it also works for future knowledge attributions.

III

There are two obvious alternatives to applying MacFarlane's assessment-context relativism to future knowledge attributions:

1) dropping the truth-condition on knowledge;

2) rejecting the idea that we *know* future contingents;

Solutions 1 involves a project that is far too ambitious to be quickly resolved here. Solution 2, on the other hand, is pretty simple: all we need to say is that we don't really know statements about the future; and that when we do claim to know them (as we often do), we misspeak; what we should really be talking of are predictions, probability, and degrees of certainty. Indeed, how can you know that something that hasn't yet happened and, in a genuinely open future, might yet not happen, will definitely happen? You don't.

I want to suggest some caution with this reply, on two grounds: firstly, the sorts of reasons for claiming that we don't really know statements about the future must not be only the same reasons supporting the indeterminacy intuition about future contingents. Because then we would end up defending the possibility of assigning the truth-value true (or false) to future contingents while at the same time rejecting the suggestion that we could then claim to know future contingents – when the only obstacle to claiming knowledge of future contingents would indeed be the truth-condition on knowledge. In short, truth must not be the only reason why we reject the claim that we know future contingents; otherwise we will have to drop the project of assigning a truth-value true (or false) to future contingents altogether.

Secondly, our reasons for rejecting knowledge of future contingents should also not just result from scepticism about induction. The worry with induction never was that I cannot know that the sun will rise tomorrow because it hasn't risen yet; but that the empirical evidence is, supposedly, not conclusive. And if it isn't conclusive, it isn't conclusive with relation to both scientific statements about the past and scientific statements about the future; with relation to both explanation and prediction.

This point can be extended to the justification condition in general: it looks as though we can be justified in believing a statement about the future as much as we are justified in believing a statement about the present or the past. The Conspirator's only evidence for believing that the President will be in the fourth vehicle tomorrow might be the very same evidence the Conspirator has for believing that the President was in the fifth vehicle the last time he travelled: a source from inside the office responsible for arranging the President's travel. So that if the Conspirator is justified in believing that the President was in the fifth vehicle the last time he travelled, then the Conspirator is justified in believing that the President will be in the fourth vehicle tomorrow.

People speak as though they know future contingents; instead of stipulating that when a person speaks that way they must be naïve, we have now offered a way to make philosophical sense of that form of speech. People speak as though they know the future; and, lo and behold, they really do.[13]

[13] Many thanks to an audience in Dublin and to the following for comments: Paul Boghossian, Niall Connolly, Douglas Edwards, Richard Hamilton, John MacFarlane, and Conor McHugh.

RETHINKING THE DEBRIEFING PARADIGM: THE RATIONALITY OF BELIEF PERSEVERANCE

David M. GODDEN

ABSTRACT: By examining particular cases of belief perseverance following the undermining of their original evidentiary grounds, this paper considers two theories of rational belief revision: foundation and coherence. Gilbert Harman has argued for coherence over foundationalism on the grounds that the foundations theory absurdly deems most of our beliefs to be not rationally held. A consequence of the unacceptability of foundationalism is that belief perseverance is rational. This paper defends the intuitive judgement that belief perseverance is irrational by offering a competing explanation of what goes on in cases like the debriefing paradigm which does not rely upon foundationalist principles but instead shows that such cases are properly viewed as instances of positive undermining of the sort described by the coherence theory.

KEYWORDS: belief perseverance, belief revision, debriefing paradigm, bounded rationality, coherence theory, foundationalism, principle of positive undermining, rationality

1. Introduction

The phenomenon of belief perseverance, which occurs when beliefs survive "the total destruction of their original evidential basis,"[1] presents at least two problems for the theory of reasoning and rationality. First is the descriptive and psychological problem of describing the nature and extent of the phenomenon, and of explaining how and why it occurs. Second is the normative and epistemological problem of whether, and to what extent, belief perseverance is rational. This paper concerns the second of these problems.

Typically belief perseverance is viewed as failure of rationality on the part of the reasoner.[2] For example, Ross, Lepper and Hubbard described the effect of their

[1] Lee Ross, Craig A. Anderson, "Shortcomings in the Attribution Process: On the Origins and Maintenance of Erroneous Social Assessments," in *Judgment under Uncertainty: Heuristics and Biases,* eds. Daniel Kahnemann, Paul Slovic, and Amos Tversky (Cambridge: Cambridge University Press, 1982), 149.

[2] Craig A. Anderson, "Belief Perseverance," in *Encyclopedia of Social Psychology*, eds. Roy Baumeister and Kathleen D. Vohs (Thousand Oaks: Sage, 2007), 109-110; Richard E Nisbett, Lee

debriefing paradigm (in which evidence demonstrating to the satisfaction of a reasoner that the original evidential basis for one of her beliefs is completely unfounded) as having "far less impact [on the reasoner's attitude to the resultant belief] than would be demanded by any logical or rational impression-formation model."[3] This intuitive view has prompted theorists of reasoning to classify belief perseverance as a cognitive bias along with phenomena like the confirmation bias,[4] the conjunction fallacy,[5] and the belief bias.[6] Indeed, it is primarily because belief perseverance is deemed to be irrational that it presents theoretical problems for accounts of human rationality and moral problems for experimenters using deception and debriefing paradigms in psychological research.

Against this view, Gilbert Harman[7] has claimed that belief perseverance is not irrational. Harman argues that the rational condemnation of belief perseverance relies on a foundationalist epistemology and theory of rational belief change. Foundationalist theories involve a principle of negative undermining which requires that subjects track all of the reasons they have for their beliefs, and make rationally appropriate adjustments to (their confidence levels in) their beliefs whenever the basing relations among them changes. Yet, Harman argues, in view of the cognitive limitations of normal human reasoners, the foundationalist theory of rational belief change is not consistent with the principles of bounded rationality. Indeed, Harman argues, on a foundationalist account most of our beliefs are not rationally held. Because of this Harman claims that the foundationalist theory of

Ross, *Human Inference: Strategies and Shortcomings of Social Judgment* (Englewood Cliffs: Prentice-Hall, 1980); Craig A. Anderson, Mark R. Lepper, and Lee Ross, "Perseverance of Social Theories: The Role of Explanation in the Persistence of Discredited Information," *Journal of Personality and Social Psychology* 39 (1980): 1037-1049.

[3] Lee Ross, Mark R. Lepper, and Michael Hubbard, "Perseverance in Self-Perception and Social Perception: Biased Attributional Processes in the Debriefing Paradigm," *Journal of Personality and Social Psychology* 32 (1975): 880.

[4] Peter C. Wason, "Reasoning," in *New Horizons in Psychology*, ed. Brian M. Foss (Harmondsworth: Penguin, 1966), 135-151; Peter C. Wason, "Reasoning About a Rule," *Quarterly Journal of Experimental Psychology* 20 (1968): 273-281.

[5] Amos Tversky and Daniel Kahneman, "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment," *Psychological Review* 90 (1983): 293-315.

[6] Jonathan St. B. T. Evans, Julie L. Barston, and Paul Pollard, "On the Conflict Between Logic and Belief in Syllogistic Reasoning," *Memory and Cognition* 11 (1983): 295-306.

[7] Gilbert Harman, *Change in View: Principles of Reasoning* (Cambridge: The MIT Press, 1986); Gilbert Harman, "Internal Critique: A Logic is not a Theory of Reasoning and a Theory of Reasoning is not a Logic," in *Handbook of the Logic of Argument and Inference: The Turn Towards the Practical*, eds. Dov M. Gabbay, Ralph H. Johnson, Hans J. Olbach, and John Woods (New York: Elsevier, 2002), 171-186.

rational belief change cannot be correct. As a corollary, belief perseverance is rational.

In place of the foundations theory Harman advocates for the coherence theory of rational belief change. The coherence theory offers an account of what goes on in cases like the debriefing paradigm which renders the belief perseverance behavior rational. This explanatory 'success' is then counted as evidence for the normative correctness of the coherence theory.

In this paper, I defend what I take to be our intuitive judgement that belief perseverance is indeed irrational. I do this by offering a competing explanation to Harman's own which classifies belief perseverance as it occurs in the debriefing paradigm as irrational without relying upon foundationalist principles. The account thereby avoids the controversial and putatively unacceptable consequence that the majority of our beliefs are not rationally held.

2. Rationality & Bounded Rationality

Reasoning (or inference) is a psychological process of reasoned change in view,[8] or belief revision, which involves "trying to improve one's overall view by adding some things and subtracting others."[9] The goal one aims at when improving one's overall view is rationality, and it is against standards of rationality that one's overall view, and the revisions made to it, are measured. It is here that epistemology and logic contribute to the theory of rational belief revision.

Historically, the normative study of rationality began with the specification of a formal system thought to embody a set of rational ideals. Judgements of rationality in individual cases were then made according to whether and how behavior satisfied the requirements of the formal system.[10] Yet, as Chater and Oaskford have observed,[11] there is something paradoxical about research in the normative qualities of human reasoning. Not only does any attempt to assess the reliability of human reasoning involve, indeed rely upon, the very processes whose reliability we are attempting to assess. But any standards we seek to provide as norms of good reasoning will themselves be products of human reasoning processes. Further, what is to be said of otherwise rational agents who regularly seem in default of the prescribed standard?

[8] Harman, "Internal Critique," 171.

[9] Gilbert Harman, "Logic, Reasoning, and Logical Form," in *Language, Mind and Brain*, eds. Thomas W. Simon and Robert J. Scholes (Hillsdale: Lawrence Erlbaum Associates, 1982), 13.

[10] Ken Manktelow, *Reasoning and Thinking* (Hove: Psychology Press, 1999), 5.

[11] Nick Chater and Mike Oaksford, "Human Rationality and the Psychology of Reasoning: Where Do We Go From Here?" *British Journal of Psychology* 92 (2001): 193-216.

Cohen argued that empirical studies cannot contribute to a demonstration of systematic irrationality in human agents.[12] Rather, Cohen argued, in order to conduct any empirical investigation into the success of human reasoning "humans have to be attributed a competence for reasoning validly, and this provides the backcloth against which we can study defects in their actual performance."[13] In the same vein, thinkers such as Dennett[14] have advanced a position which Stich called the *argument from the inevitable rationality of believers*.[15] On Stich's reconstruction, Dennett does not hold that people must be rational, but that "people must be rational *if they can usefully be viewed as having any beliefs at all*," or as Stich puts it "intentional descriptions and rationality come in the same package."[16] Against these views Stich has argued that *a priori* arguments seeking to show that human irrationality cannot be empirically demonstrated are miscast.[17]

My own view is roughly that of Peirce, as set out in "The fixation of belief," where he maintained that we are "in the main logical animals, but we are not perfectly so."[18] Perhaps another way to state this type of position, more in line with the notion of bounded rationality and evolutionary epistemology, can be found in Nisbett and Ross's thesis that "people's inferential strategies are well adapted to deal with a wide range of problems, but that these same strategies become a liability when applied beyond that range."[19]

2.1. Bounded Rationality

Yet, there is something to Cohen's view that our rational norms must be competence norms. If a normative theory of reasoning is to be prescriptive over our actual inferential practices, then it seems as though we must be able to follow the prescriptions made by the theory. So, a general problem for theorists seeking to provide a normative theory of reasoning stems from our nature as rational agents with a finite cognitive endowment.

A cardinal example of this is the idea of deductive closure: that the beliefs of a perfectly rational agent should be closed under the laws of deduction. A

[12] Jonathan L. Cohen, "Can Human Irrationality Be Experimentally Demonstrated?" *Behavioral and Brain Sciences* 4 (1981): 317-370.

[13] Cohen, "Can Human Irrationality," 317.

[14] Daniel Dennett, *Brainstorms* (Montgomery: Bradford Books, 1978).

[15] Stephen Stich, "Could Man Be An Irrational Animal?" *Synthese* 64 (1985): 120.

[16] Stich, "Could Man Be," *Synthese* 64 (1985): 121.

[17] Stich, "Could Man Be," Stephen Stich, *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation* (Cambridge: The MIT Press, 1990).

[18] Charles Sanders Peirce, "The Fixation of Belief," *Popular Science Monthly* 12 (1977): 1-15.

[19] Nisbett and Ross, *Human Inference*, xii.

consequence of this principle is that agents should believe all of the logical consequences of their present beliefs. Yet, even a single belief, *p*, generates an infinitude of logical consequences through the law of disjunction introduction ($p \vdash p \vee q$). Thus, any single belief, *p*, would generate the following infinite series of logical consequences: $p \vee q$, $(p \vee q) \vee r$, $[(p \vee q) \vee r] \vee s$, …, etc. As Harman rightly points out, not only is it impossible for a finite cognitive agent to follow through on all of the consequences of her present beliefs, it is often neither practical nor prudent to even begin to do so.[20] With this in mind, Harman introduces the principle of *clutter avoidance* – that "one should not clutter one's mind with trivialities" – as an example of the type of principle that belongs in a properly articulated theory of rational belief change. [21]

It is not just that the principles of ideal rationality do not apply to agents whose powers of reasoning are finite; they also do not apply to rational agents whose judgement is fallible. Take, for example, the principle of consistency: that all of our beliefs should be consistent with one another. Using an example like the preface paradox,[22] Harman observes not only is it not possible to attain perfect consistency among all of our beliefs, for fallible judges it is not rational to have perfectly consistent beliefs. He writes:

> a rational fallible person ought to believe that at least one of his or her beliefs is false. But then not all of his or her beliefs can be true, since, if all of the other beliefs are true, this last one will be false. So in this sense a rational person's beliefs are inconsistent. It can be proved they cannot be all true together.[23]

Considering this type of case, Pinto argues (i) that so long as one does not use contradictory premises to make inferences, that the rational fault of having inconsistent beliefs can be no more serious than that of having a false belief,[24] and (ii) that unless one has a particular reason to suspect one of the beliefs involved in the inconsistency is dubious, it is not rational to give any of them up upon discovering an inconsistency.[25] Thus, it should be recognized at the outset that the ideals of perfect rationality simply do not apply in any direct fashion to cognitively

[20] Harman, *Change in View*, 12.

[21] Harman, *Change in View*, 12.

[22] See David C. Makinson, "The Paradox of the Preface," *Analysis* 25 (1965): 205-207.

[23] Gilbert Harman, "Logic and Reasoning," *Synthese* 60 (1984): 107.

[24] Robert C. Pinto, "Inconsistency, Rationality and Relativism," in *Argument, Inference and Dialectic: Collected Papers on Informal Logic*, eds. Robert C. Pinto and Hans V. Hansen (Dordrecht: Kluwer, 2001), 46-51.

[25] Pinto, "Inconsistency, Rationality," 51-53.

finite rational agents, and cannot provide the sole basis of rational norms for human reasoners.

Theories of bounded rationality, which take into account the finite limitations of human cognitive ability, provide alternatives to accounts of perfect or idealized rationality which are based solely on abstract logical systems. According to bounded rationality, the prescriptiveness (or normative standing) of a set of norms derives not only from its connection to an abstract or formal standard (e.g. deductive closure, absence of contradiction, etc.) but also from the fact that such standards can be attained in principle by human reasoners.[26] That is, it is within our capacity to reason in accordance with the norms: that we ought to reason in a particular way implies that we can reason in that way.

2.2. Cognitive Biases in Reasoning

Despite the initial plausibility of the idea of bounded rationality, a question quickly emerges: to what degree should actual performance be taken as the measure of competence? Should the untutored behavior of normal cognitive agents serve as a basis for prescriptive theories of rationality? Problematically, there are many well known cognitive biases – fallacies of reasoning if you will – which typify human cognitive habits. Evans writes that "A 'bias' is usually defined as systematic attention to some logically irrelevant features of the task, or systematic neglect of a relevant feature."[27] One such characteristic failure of reasoning, which has come to be called the phenomenon of belief perseverance, was described by Francis Bacon in the *New Organon*:

> The human understanding when it has once adopted an opinion draws all things else to support and agree with it. And though there be a greater number and weight of instances to be found on the other side, yet these it either neglects and despises, or else by some distinction sets aside and rejects, in order that by this great and pernicious predetermination the authority of its former conclusion may remain inviolate.[28]

It is this with cognitive bias that I am presently concerned, and I will consider it as it has been studied in a series of experiments known as the debriefing paradigm.

[26] Cf. Jonathan Baron, *Rationality and Intelligence* (Cambridge: Cambridge University Press, 1985) for a distinction between normative and prescriptive theories.

[27] Jonathan St. B.T. Evans, "Bias and Rationality," in *Rationality: Psychological and Philosophical Perspectives*, eds. Ken I. Manktelow and David E. Over (New York: Routledge, 1993), 16.

[28] Francis Bacon, *New Organon* (1620), quoted in Nisbett and Ross, *Human Inference*, 167.

3. The Debriefing Paradigm

3.1. The Paradigm Described

While familiar to psychologists, a recapitulation of the debriefing paradigm might still be in order. In general the debriefing paradigm works as follows: it is designed to prompt a participant reasoner (either an actor or an observer) to form a belief (e.g., their success at a given task) (actors form such beliefs about themselves, observers about actors), under controlled circumstances on the basis of certain evidence (feedback given during the performance of the task) which is subsequently undermined (the feedback is shown to be false); participants then report on the status of the resultant and related beliefs (e.g., actual success on given task, on future tasks, and in general on related tasks). The result is that beliefs so formed often survive the complete undermining of the evidence on the basis of which they were formed.

A standard experimental paradigm[29] involves getting participants to distinguish between supposedly authentic and fake suicide notes. During the task, participants are given false feedback which ranks their success as either above average, average, or below average. Following the task each participant is completely debriefed: the false and predetermined nature of the feedback is thoroughly explained. On the (standard) outcome debriefing condition, participants are told that the feedback was contrived and not in any way linked to their actual performance; they are shown the experimenter's instructions specifying the details of the feedback to be given and assigning them to the success, average or failure group.[30] Finally, as part of an ostensibly unrelated questionnaire, participants are asked to estimate their actual performance on the task they completed, and their prospects for future success both in similar tasks and in general.

The result is that even following debriefing, participants assigned to the success group (and their observers) ranked their abilities more highly than those assigned to the average or failing groups. Ross, Lepper and Hubbard summarized the results this way:

> even after debriefing procedures that led subjects to say that they understood the decisive invalidation of initial test results, the subjects continued to assess their performances and abilities as if these test results still possessed some validity.[31]

[29] Ross, Lepper, and Hubbard, "Perseverance in Self-Perception."

[30] cf. Ross, Lepper, and Hubbard, "Perseverance in Self-Perception," 883, 885; Cathy McFarland, Adeline Cheam, and Roger Buehler, "The Perseverance Effect in the Debriefing Paradigm: Replication and Extension," *Journal of Experimental Social Psychology* 43 (2007): 234, 235-6.

[31] Ross, Lepper, and Hubbard, "Perseverance in Self-Perception," 884.

That is, the inferred belief perseveres "[even] when a person discovers that the entire evidence base for the initial [inferred] judgement is not merely biased or tainted but is completely without value."[32] Despite the fact that most reasoners seem to behave in this way, the intuitive judgement of most theorists is that this behavior is irrational, and that reasoners ought to abandon, or at least revise their confidence in, the belief about their abilities based on the false feedback. Indeed, the results of the debriefing paradigm seem paradoxical because the reasoner, through the process of the experiment, recognizes the defeat of a belief (about the authenticity of the feedback) she has, but she does not recognize this as the defeat of a set of reasons on the basis of which she adopted an additional belief (about her ability).

Subsequent work with the debriefing paradigm[33] has not only confirmed the results of the initial studies, but has extended them in several directions. Perhaps most remarkable is a study by Wegner, Coulton and Wenzlaff[34] which demonstrated that even when the falsity of the feedback was made apparent to participants *prior* to their receiving it (i.e., briefing rather than debriefing), participants still treated the (false) information as though it were true and not completely undermined. "In sum," they wrote, "briefing and debriefing had essentially equivalent effects, leading neither actors nor observers to forsake the feedback as a cue to the actor's performance."[35]

Empirically, two debriefing techniques have been found to effectively the counter-act the perseverance effect. First, Ross, Lepper and Hubbard[36] reported that process debriefing, whereby in addition to a normal, outcome debriefing, participants are also informed about the phenomenon of belief perseverance and how it can occur, and then warned of its potential personal relevance to them in the context of the experiment, significantly reduced and often effectively eliminated the perseverance effect. Second, McFarland, Cheam and Buehler[37] reported that a revised outcome debriefing which, in addition to a normal outcome debriefing, informed participants of the invalidity of the entire test, eliminates the perseverance effect just as well as process debriefing. Revised outcome debriefing seeks to make manifest to the participant not merely that the information given in feedback is

[32] Nisbett and Ross, *Human Inference*, 176.

[33] e.g., McFarland, Cheam, and Buehler, "The Perseverance Effect in the Debriefing Paradigm."

[34] Daniel M. Wegner, Gary F. Coulton, and Richard Wenzlaff, "The Transparency of Denial: Briefing in the Debriefing Paradigm," *Journal of Personality and Social Psychology* 49 (1985): 338-346.

[35] Wegner, Coulton, and Wenzlaff, "The Transparency of Denial," 342.

[36] Ross, Lepper, and Hubbard, "Perseverance in Self-Perception."

[37] McFarland, Cheam, and Buehler, "The Perseverance Effect in the Debriefing Paradigm," 235-236.

false, but that the task itself is entirely fake, i.e., that the source of the information has no potential whatsoever to be reliable.[38]

3.2. Explaining our Intuitions Concerning the Rationality of Belief Perseverance

Perhaps the first step in appreciating the explanatory paradoxes raised by belief perseverance is to get a handle on the intuitions theorists tend to have concerning its irrationality. The reason behind this type of intuition seems to be something like this: wholly unjustified beliefs ought to be abandoned. As Ross, Lepper and Hubbard put it: "With no pertinent information remaining, ... the perceiver's assessment [should] return to its starting point as any logical or rational impression-formation model would demand."[39] Indeed, this intuition seems to be a corollary of some other, more general, principles of rationality.

The first of these might be the principle of *evidence proportionalism* which has been defined by Engel (following Hume[40]) as follows: "In general a belief is rational if it is proportioned to the degree of evidence that one has for its truth."[41] The rationality of belief is explained, at least in part, through a kind of evidence proportionalism. Thus, beliefs without any evidential support ought to be abandoned. Specifically, the persevering belief, having been deprived of the only supporting evidence on the basis of which it was adopted, seems to be completely unsupported, thereby rationally requiring its abandonment.

A second general principle which perhaps underlies our intuitions concerning the rationality of belief perseverance is the *principle of commutativity*, which Nisbett and Ross formulate as follows: "the net effect of evidence A followed by evidence B must [i.e., ought to] be the same as for evidence B followed by evidence A."[42] Roughly, the order in which we receive information ought not to affect its overall significance or evidential force. Bringing this principle to bear on belief perseverance treats it as a limiting case of the primacy effect whereby newly received information is synthesized in such a way as to represent is as consistent

[38] Seemingly, in Ross, Lepper, and Hubbard's initial (1975) experiment the task was genuine: there were authentic as well as inauthentic suicide notes in each pair, and thus a correct answer was possible on any trial. By the time McFarland, Cheam, and Buhler (2007) repeat the study, it seems that all the notes are inauthentic, and thus that the test itself was a fabrication, with no possibility of measuring what it purported to measure.

[39] Ross, Lepper, and Hubbard, "Perseverance in Self-Perception," 881.

[40] David Hume, *Enquiries Concerning Human Understanding and Concerning the Principles of Morals* (1777) (Oxford: Clarendon Press, 1975), X.i.87; 110.

[41] Pascal Engel, "Introduction: The Varieties of Belief and Acceptance," in *Believing and Accepting*, ed. Pascal Engel (Dordrecht: Kluwer, 2000), 3.

[42] Nisbett and Ross, *Human Inference*, 169.

and coherent with previously received information. That is, reasoners are maximally conservative when revising existing beliefs, tending instead to interpret the significance of new information in light of already accepted information. Recognizing that the order in which we receive information is normally irrelevant to its evidential import, the principal of commutativity condemns the perseverance effect as irrational.

3.3. Explaining the Results of the Debriefing Paradigm

A variety of theories exist that contribute to a psychological explanation of the phenomenon of belief perseverance. Anderson[43] discusses three psychological processes that contribute to such an explanation: the availability heuristic (where only memorable confirming or disconfirming cases are considered); illusory correlation (where more confirming cases and fewer disconfirming ones are remembered than actually exist); and data distortions (where "confirming cases are inadvertently created and disconfirming cases are ignored").

The operation of these cognitive biases can easily be appreciated when imagining the reasoning process of participants in the debriefing paradigm. Nisbett and Ross imagine participants, having adopted the target belief, searching around for confirming evidence among their existing beliefs. For example, a participant who is given (false) feedback that she is a successful discriminator of genuine versus faked suicide notes could take her "reasonably good performance in her abnormal psychology course, her ability to make new friends easily, and her increasing sense of confidence and assurance as she progressed in the … task" as "further evidence" of her discriminatory powers. Similarly, a participant given negative feedback "might note her difficulty in imagining herself as lonesome or alienated, her mediocre performance in her social problems course, and her increasing sense of confusion and hesitation as she progressed in the … task" as further evidence of her own unreliability.[44] There is adequate confirming evidence no matter which way things go. This scenario can occasion an important hypothesis, for it promises to show something important about how humans learn and synthesize information. When we acquire a new piece of information we synthesize it with our existing beliefs by finding ways that it can serve as a premise or conclusion from our existing beliefs. As we find new ways that the belief can be a conclusion of our existing beliefs, it becomes further entrenched in our overall web of belief.

[43] Anderson, "Belief Perseverance," 110.

[44] Nisbett and Ross, *Human Inference*, 181.

Another explanatory hypothesis, first suggested by Ross, Lepper and Hubbard[45] and later by Anderson, Lepper and Ross[46] is that the participant finds *explanations* of the result from among their existing beliefs. That is, the new piece of information is treated as an explanandum, and reasoners search around for an explanans among their existing beliefs. Here, the participant's beliefs about her performance in her abnormal psychology or social problems course does not serve as evidence for her belief about her 'performance' (as described by the false feedback) in the experimental task, but rather serves to causally explain her 'performance' in the experiment. Since the explanans already occurs among her existing beliefs (i.e., the usual evidential order for a causal explanation is reversed), this further allows the participant to accept the belief because, having explained it, they are entitled to expect the result, or better understand why it happened. Problematically, though, as Anderson, Lepper and Ross observe, "[o]nce a causal account has been generated, it will continue to imply the likelihood of the 'explained' state of affairs even after the original basis for believing in that state of affairs has been eliminated."[47]

Evolutionary explanations tend to claim that the acquisition of new information is costly. Therefore, once a piece of information has been acquired a cognitively economical strategy is to be as conservative as possible when it comes to revising or abandoning one's doxastic attitude to that information.

Having surveyed some of the explanations of the phenomenon of belief perseverance arising from the debriefing paradigm, the task of more closely scrutinizing the rationality of the behavior remains.

4. Foundations and Coherence

In *Change in View* Harman examines the relationship between justification and belief revision, considering two theories of justification: the foundations theory and the coherence theory. As a theory of justification, each theory serves as a model of ideal belief revision.[48] In distinguishing the two theories, Harman writes: "[t]he key issue is whether one needs to keep track of one's original justifications for beliefs," foundationalists say "yes" and coherentists say "no."[49] Thus, Harman writes, "the theories are most easily distinguished by the conflicting advice they occasionally give concerning whether one should *give up* a belief P … when P's original

[45] Ross, Lepper, and Hubbard, "Perseverance in Self-Perception," 890.

[46] Anderson, Lepper, and Ross, "Perseverance of Social Theories."

[47] Anderson, Lepper, and Ross, "Perseverance of Social Theories," 1038.

[48] Harman, *Change in View*, 29.

[49] Harman, *Change in View*, 29.

justification has to be abandoned."[50] As theories of justified belief, foundations and coherence can be roughly characterized as follows:

- *Foundations theory of justified belief*: one is justified in continuing to believe something only if one has a special reason to continue to accept that belief;
- *Coherence theory of justified belief*: one is justified in continuing to believe something as long as one has no special reason to stop believing it.[51]

Corresponding to these two theories of justified belief, are two corollary positions concerning belief revision.[52] Deriving from the foundations theory is the

> *Principle of negative undermining*: One should stop believing *P* whenever one does not associate one's belief in *P* with an adequate justification (either intrinsic or extrinsic),[53]

while the coherence theory gives us the

> *Principle of positive undermining*: One should stop believing *P* whenever one positively believes one's reasons for believing *P* are no good.[54]

As Harman construes it, the foundations theory holds that instances of belief perseverance arising from the debriefing paradigm violate the principle of negative undermining and are therefore irrational, while the coherence theory licenses this behavior as being consistent with the principle of positive undermining and therefore rational.

As Harman observes, when we consider actual cases of belief perseverance it becomes clear that there are problems with the descriptive accuracy of the foundations theory. It is found that people retain beliefs even after the positive refutation of all the evidence that was originally provided in support of the belief. To explain this phenomenon, Harman suggests that

[50] Harman, *Change in View*, 30.

[51] Harman, *Change in View*, 36.

[52] Harman, *Change in View*, 39.

[53] For Harman (*Change in View*, 30-31), the justification of basic beliefs is intrinsic while derived beliefs – beliefs whose justification relies on other beliefs – have extrinsic justifications.

[54] In considering the debriefing paradigm, Goldman proposes a similar rule. Having rejected a rule which instructs one to "Revise all ...beliefs that have been undermined by new evidence," and even a "rule system [which] would *oblige* a cognizer *continually* to search for old beliefs in LTM [long term memory] that might be weeded out in light of new evidence," he proposes a rule which prescribes "if one activates an old belief in q, and if one (actively) believes that this belief wholly stems from now abandoned evidence, then one is required to abandon q" (Alvin I. Goldman, *Epistemology and Cognition* (Cambridge: Harvard University Press, 1986), 220-21).

> what the debriefing studies show is that people simply do not keep track of the justification relations among their beliefs. They continue to believe things after the evidence for them has been discredited because they do not realize what they are doing. They do not understand that the discredited evidence was the *sole* reason why they believe as they do.[55]

Yet, Harman argues that the foundations theory not only fails descriptively; it also fails as a prescriptive theory. Harman argues that "[P1] the [foundations] theory implies that people are unjustified in almost all their beliefs. [And P2] This is an absurd result." The reasons Harman gives for P1 are [Pi] that the foundations theory requires that agents keep track of their reasons for their beliefs in order for their beliefs to be rationally justified and, [Pii] as evidenced by studies like the debriefing paradigm, "people rarely keep track of their reasons [for their beliefs]."[56] Indeed the computational costs for any cognitively finite agent attempting to actively track all of the evidentiary relations on the basis of which it (comes to) hold(s) its beliefs is intractable.[57] Since any normative theory which classifies the majority of our beliefs as irrational violates Cohen's basic principle of judging instances of irrational performance against a backcloth of rational competence, the foundations theory cannot be accepted as providing the normative standards against which the rationality of our inferential behavior should be assessed. Therefore, Harman concludes, "The foundations theory turns out not to be a plausible normative [i.e., prescriptive] theory [of rational belief change] after all."[58]

By our rational intuitions alone, we tend to judge the belief perseverance behavior of participants in the debriefing paradigm as irrational. Yet according to Harman, these rational intuitions presuppose a foundations approach to reasoned belief revision which is unacceptable. Further, since on the coherence theory the majority of our actual beliefs are rationally held, it is a better prescriptive theory than foundationalism. As a corollary, belief perseverance behavior also turns out to

[55] Harman, *Change in View*, 38. This seems to be a point which Ross, Lepper and Hubbard themselves accepted. They wrote: "We propose that first impressions may not only be enhanced by subsequent biases in coding but may ultimately be *sustained* through such biases. The perceiver, we contend, typically does not reinterpret or reattribute impression-relevant data when the basis for his original coding bias is discredited; *once coded, the evidence becomes autonomous from the coding scheme, and its impact ceases to depend upon the validity of that schema*" (Ross, Lepper, and Hubbard, "Perseverance in Self-Perception," 889, emphasis added).

[56] Harman, *Change in View*, 39.

[57] Mike Oaksford and Nick Chater, "Reasoning Theories and Bounded Rationality," in *Rationality: Psychological and Philosophical Perspectives*, 31-60; Mike Oaksford and Nick Chater, "Theories of Reasoning and the Computational Explanation of Everyday Inference," *Thinking and Reasoning* 1 (1995): 121-152.

[58] Harman, *Change in View*, 39.

be rational. Basically, Harman's view (as I read it) is that our intuitions about the rationality of belief perseverance and the debriefing paradigm are irrational not the behavior of reasoners in these cases.

5. Debriefing: Why Failure to Track Reasons is not the Problem

Against Harman, I hold that our intuitions about the irrationality of belief perseverance in cases like the debriefing paradigm are correct. In responding to his position I do not propose an argument in favor of the foundations theory of rational belief revision. Rather, I deny that our intuitive judgements of the irrationality of belief perseverance presuppose the foundations theory. Instead, I provide an account of what is going on in the debriefing paradigm which confirms our intuitions about the irrationality of the results of the paradigm without requiring that reasoners track their reasons for their beliefs.

On Harman's account, the failure of the reasoner is not one of rationality but one of memory. She has simply forgotten that the defeated reasons were the reasons on the basis of which she initially adopted a belief and it is because of this that she does not draw the connection between the defeat of those reasons and the acceptability of the belief. On this picture, even the failure to actively track her reasons – regardless of whether those reasons become defeated or overridden – is enough to violate the principle of negative undermining thereby requiring abandoning the belief. By contrast, the principle of positive undermining is never violated since the reasoner never realizes that her *reasons* have been defeated. That is, she never comes to the positive belief that her reasons for her belief are no good: not because she does not recognize the defeat of certain claims (the reasons themselves) but because she fails to connect those claims with the belief for which they served as reasons.

But the question remains: is this failure to make the connection between the defeat of the reasons and the acceptability of the belief properly explained as a failure of memory? I maintain that the results of the debriefing paradigm, and others like it, are not cases of negative undermining and are not properly explained or justified as a failure of memory. Instead, these are cases of positive undermining, where reasoners fail to recognize the evidentiary significance of new information available to them. Even if the reasoner has not tracked (e.g., by forgetting) her initial reasons for adopting the belief, this neither explains nor excuses her failure to examine the acceptability of the belief following the defeat of those reasons. On my account, the failure in such cases is not an understandable failure of memory but a rationally reprehensible failure to see the immediate consequences of new information. This failure to recognize the immediate consequences of new information is best understood not as a failure of memory but as a failure of understanding.

It is important to note that accounts involving both the foundations theory and the coherence theory presuppose that the reasoner has not arrived at any *new* or *alternate* reasons for the target belief. If it is supposed that the reasoner possesses any new or alternate reasons supporting the belief, then the case ceases to be problematic: we would not intuitively deem it irrational that she would continue to hold the belief, and neither would either of the two theories.

5.1. Understanding and Drawing the Right Inferences

In the debriefing paradigm, the participant reasoner drew an inference regarding her abilities on the basis of the information given to her as feedback during the experiment. We might call the connection in her mind between the feedback and the conclusion she drew therefrom her cognitive warrant.[59] That is to say that part of the significance to the reasoner of the information given in the feedback is that it yielded certain consequences. Yet when, during the debriefing, the reasoner recognized that this same information was defeated, she did not thereby recognize that the conclusion she drew therefrom might be undermined. That is, she failed to appreciate the immediate consequences of new information she had come to accept. There is no need for her to have tracked or remembered the initial inference she drew. Rather, the only thing that is required of her is that she recognize the significance of the information immediately before her. On this account the failure is not one of memory but of understanding.

Part of what it is to understand a claim is to understand how it connects inferentially to other claims. That is, part of what it is to understand a claim is to know what other claims it could be concluded from, and what other claims it could serve as a premise for. This view of meaning and understanding Brandom calls *inferentialism*[60]

> Understanding or grasping a propositional content is here presented not as the turning on of a Cartesian light, but as a practical mastery of a certain kind of inferentially articulated doing: responding differentially according to the circumstances of proper application of a concept, and distinguishing the proper inferential consequences of such application. … Thinking clearly is on this inferentialist rendering a matter of knowing what one is committing oneself to by a certain claim, and what would entitle one to that commitment. … Failure to

[59] See the next section for an explanation of the notion of a cognitive warrant.

[60] With Brandom, I see this as part of a pragmatic view of meaning which holds that meaning is explained in terms of use, and to understand a linguistic expression is to know how it is correctly used. One of the ways that we use statements is as premises and conclusions in inferences. Thus, part of what it is to understand a claim is to know what inferences can be correctly made involving them.

> grasp either of these components is failure to grasp the inferential commitment that use of the concept involves, and so failure to grasp its conceptual content.[61]

While Brandom here talks in terms of understanding propositions, elsewhere he specifically links this inferentialist account of understanding to the contents of beliefs, writing: "Understanding the content of a speech act or belief is being able to accord the performance of that speech act or the acquisition of that belief the proper practical significance – knowing how it would change the score [of commitments and entitlements] in various contexts."[62] So, to correctly understand the meaning of a claim is to understand its inferential significance. To the degree to which we fail to appreciate the consequences of a claim (the commitments it puts upon us), we fail to understand it. Similarly, to the degree to which we fail to appreciate what a claim is a consequence of (what would entitle us to it), we fail to understand it.

In the case of the debriefing paradigm, the participant reasoner fails to appreciate the commitments put upon her by her accepting during debriefing that the information given as feedback is indeed false. What makes cases like this so remarkable is not that an individual has forgotten her reasons for adopting a belief. Rather, it is that on one occasion she recognized some information as having a certain significance by immediately making certain inferences on its basis. Yet, on another occasion she fails to recognize the significance of that same information by failing to make the relevant inferences. Seen in this way, the reasoner fails to appreciate the meaning of the information, or at least treats the information differently from one occasion to the next. Rather than forgetting their reasons for a belief which they may not be attending to, reasoners in the debriefing paradigm fail to appreciate the significance of information immediately present to them and thereby misunderstand it.

[61] Robert B. Brandom, *Articulating Reasons: An Introduction to Inferentialism* (Cambridge: Harvard University Press, 2000), 63-64.

[62] Brandom, *Articulating Reasons*, 165-166. The passage preceding the quoted sentence reads as follows: "One can pick out what is *propositionally* contentful to begin with as whatever can serve both as a premise and as a conclusion in *inference* – what can be offered as, and itself stand in need of, *reasons*. Understanding or grasping such a propositional content is a kind of know-how – practical mastery of the game of giving and asking for reasons, being able to tell what is a reason for what, distinguish good reasons from bad. To play such a game is to keep *score* on what various interlocutors are committed and entitled to. Understanding the content of a speech act or belief is being able to accord the performance of that speech act or the acquisition of that belief the proper practical significance – knowing how it would change the score in various contexts" (Brandom, *Articulating Reasons*, 165-166).

Two situational facts about the experimental method of the debriefing paradigm make this result especially conspicuous. First is the fact that the inferential path from the feedback to the target belief was not complicated but was an immediate inference for the participant reasoner. No other information was required in making the inference in the initial instance. Second is the fact that the time between the initial feedback and its subsequent undermining is not especially long.[63] *Ex hypothesi*, not only was there no occasion for the discovery of *new* evidence for the target belief but there was no occasion for the re-evaluation or displacement of the cognitive warrant initially relied upon.

Viewed in this way, the prescribed result of the debriefing paradigm does not seem nearly as cognitively onerous as Harman's account makes it out to be. Yet the question remains as to whether the situation of the debriefing paradigm is properly interpreted as an instance of (unrecognized) positive undermining. To answer this question we must consider the notion of a cognitive warrant a little more closely.

5.2. Cognitive Warrants and Positive Undermining

Harman's *principle of positive undermining* states: "one should stop believing *P* whenever one positively believes one's reasons for believing *P* are no good."[64] The question is, how are external judges to determine when an individual reasoner judges – or ought to judge – that her reasons for believing something are good or no good? To make this determination solely on the basis of whether the reasoner actually comes to adopt or abandon some belief cannot be accepted as a satisfactory. The problem with this method is that it presumes that the reasoner is never mistaken or irrational. Yet since cases like belief perseverance raise questions about the rationality of individual reasoners in certain situations, we should not allow this question to be begged.

Instead, what is needed is a way of determining when the principle of positive undermining has been satisfied, even if a reasoner has failed to recognize this or to act on it appropriately. I suggest that a neutral way of attempting to determine when a reasoner (ought to) positively believe(s) that her reasons for believing something are (no) good is to invoke the idea of a cognitive warrant.

[63] Ross, Lepper, and Hubbard ("Perseverance in Self-Perception") initially tested for two intervals, 5-minutes and 25-minutes between the conclusion of the briefing-phase and debriefing. (During this time participants are not exposed to any new information.) Finding no statistically significant difference between these conditions, they opted for the shorter interval. Similarly, McFarland, Cheam and Buehler use a delay interval of "a few minutes" (McFarland, Cheam, and Buehler, "The Perseverance Effect in the Debriefing Paradigm," 235).

[64] Harman, *Change in View*, 39.

A warrant is like an inference ticket: it is a rule that licenses or underwrites a move to infer some claim on the basis of other claims. A cognitive warrant is the warrant that a reasoner actually uses or actually draws upon in making an inference on some particular occasion. Cognitive warrants can be understood as something like the "habits of mind" Peirce described in "The Fixation of Belief" as "that which determines us, from given premises, to draw one inference rather than another" and which Peirce there called a *guiding principle* of inference.[65]

A cognitive warrant is psychological in nature and need not be explicitly formulated in the mind of the reasoner in order to be operative. A cognitive warrant might have no logical or epistemological merit whatsoever instead being based on nothing more than a psychological association in the mind of a reasoner. But even as such, it has psychological force for that individual reasoner. Further, a reasoner need not articulate the warrant to herself when relying upon it in her reasoning; indeed she need not be aware of it at all. Cognitive warrants are markers of consequences that some reasoner finds immediately apparent when presented with certain information. It is the job of epistemologists and psychologists of reasoning to make these cognitive warrants explicit in an attempt to explain and evaluate processes of reasoning.

Importantly, Harman's account of reasoning relies on notions very similar to these cognitive warrants or guiding principles of inference. Harman assumes that "one has certain basic dispositions to take some propositions immediately to imply other propositions and to take some propositions as immediately inconsistent with each other,"[66] thereby introducing the notions of immediate implication and immediate inconsistency which are defined internally to the psychology of individual reasoners. These notions are meant to replace the purely logical notions of implication and inconsistency in formulating prescriptive norms for reasoners. For example, Harman's *Principle of Immediate Implication* states: "That *P* is immediately implied by things one believes can be a reason to believe *P*."[67] In the debriefing paradigm, it would seem that, even on Harman's account, the target belief concerning a reasoner's task-related abilities is an immediate implication of the feedback given during the experiment. Whatever the link is in the mind of the reasoner between the feedback and the target belief we can call her cognitive warrant.

[65] Peirce, "The Fixation of Belief."

[66] Harman, *Change in View*, 19.

[67] Harman, *Change in View*, 21.

The truth of such guiding principles of inference, according to Peirce, "depends on the validity of the inferences which the habit determines."[68] Thus, cognitive warrants can be objectively evaluated against some external standard of rationality according to whether they are (generally) truth-preserving or reliable. But, more importantly for the purposes of this argument, the reasoning of individuals can also be rationally evaluated by an internal standard according to whether the agent consistently applies the cognitive warrants they actually (though perhaps tacitly) accept from one occasion to the next.

That a reasoner relies on, or acts in accordance with, a cognitive warrant on some occasion is an indication that she finds it to mark a relationship of immediate implication between two (sets of) beliefs. That is, at some (perhaps unconscious) level she finds the premissory beliefs to be good reasons for the conclusion-belief. Because of this, we should be able to use the notion of cognitive warrants to determine when the *principle of positive undermining* ought to apply. Namely, we can say that a reasoner, S, ought to believe that her reasons for believing that *P* are no good whenever a set of beliefs that immediately imply *P* (for S) have been manifestly (to S) shown to be unacceptable. That is, a belief, *P*, has been positively undermined whenever a cognitive warrant having *P* as its conclusion has been positively undermined. Yet, this is exactly what happens in the case of the debriefing paradigm: participants are shown that their reasons – reasons which immediately implied some target belief – are no good.

To better appreciate this, consider the way that Ross, Lepper and Hubbard described standard outcome debriefing:

> The experimenter explained that the subject's success or failure had been randomly determined prior to her arrival. He emphasized that the subject's score had not been dependent on her performance and that it provided absolutely no information about her actual performance.[69]

Similarly, McFarland, Cheam and Buehler describe (standard) outcome debriefing as follows:

> Participants in the *standard outcome debriefing* condition were informed that their score was a fake score that had been randomly assigned to them prior to their arrival. Additionally, they were shown a "random assignment schedule," and the experimenter emphasized that the score contained absolutely no information about the participant's actual performance or underlying abilities.[70]

[68] Peirce, "The Fixation of Belief."

[69] Ross, Lepper, and Hubbard "Perseverance in Self-Perception," 885.

[70] McFarland, Cheam, and Buehler, "The Perseverance Effect in the Debriefing Paradigm," 235.

The emphasis offered in debriefing has the effect of reminding the participant of the inference she had drawn only minutes ago on the basis of the now defeated information. In doing so, it makes manifest to the participant that the cognitive warrant she relied on just previously is entirely defeated and her inference thereby undermined.

Seen in this way, the debriefing paradigm is a situation of positive rather than negative undermining. Recognizing that the undermining occurs does not require tracking any reasons, but only requires that the reasoner recognizes the immediate implications of the information presented, indeed emphasized, to her in debriefing. Her failure to see that the target belief has been undermined is evidence not that she is forgetting what her reasons for that belief were, but that she is misunderstanding (or ignoring the probative significance of) the information immediately before her by failing to apply the same cognitive warrant on one occasion that she had (and was reminded she had) applied only minutes previously. And, in accordance with our intuitions, this behavior is irrational.

The problem, I suggest, with Harman's approach to assessing the rationality of belief perseverance in debriefing is not his advocacy of a coherentist principle of positive undermining over a foundationalist principle of negative undermining. Rather, the problem is that he psychologizes the criteria for determining when positive undermining has occurred. Positive undermining occurs when one comes to believe that one's reasons for believing some claim are no good. Yet, as a criteria for determining when this occurs, Harman seems to invoke his *Immediate Inconsistency Principle*: "Immediate logical inconsistency in one's view can be a reason to modify one's view."[71] It would seem that Harman takes positive undermining to have occurred only if an agent positively forms a belief that the defeat of her reasons (during debriefing) is inconsistent with her acceptance of the target belief. In other words, positive undermining has not occurred unless the reasoner *recognizes* not merely that her *reasons* for P are defeated, but also that this is somehow inconsistent with continuing to hold that P. But to use this as the criterion for positive undermining is to presuppose that the reasoner is always rational, and it is precisely the rationality of the reasoner that is at issue. Positive undermining can occur, and yet the reasoner can fail to recognize it.

This problem is overcome, I suggest, by the proposed method. The cognitive warrant method is neutral concerning objectively correct rational norms because it does not impose any external normative standard of reasoning on the reasoner, instead relying on her own inferential habits to determine which inferences she takes to be good ones. At the same time, it allows third-party judges to hold

[71] Harman, *Change in View*, 22.

reasoners accountable to their own putative standards, by insisting that they give the same information the same significance from one occasion to the next by applying the same cognitive warrants to it as they have in the immediate past. On this model it is possible for a reasoner to fail to recognize that a belief has been positively undermined, and it is possible for a third-party to make a judgement about when this has occurred.

This is not to say that reasoners cannot abandon or change their conscious attitudes towards their cognitive warrants just as they can their beliefs. It does, though, allow theorists of reasoning a way of making explicit those rules which characterize a reasoner's inferential habits, and thereby of talking about when it is rational for reasoners to change those inferential habits. Because of the experimental conditions of the debriefing paradigm and the temporal proximity from feedback to debriefing it is not reasonable to suppose that the cognitive warrant relied upon in the feedback stage should have been abandoned or re-evaluated prior to debriefing.

5.3. An Objection to the Proposed Account

An important objection to the account just described is that the defeat of antecedent information occasioning the drawing of a specific inference need not undermine the conclusion drawn in that inference.

Let us represent the cognitive warrant used by the participant reasoner, S, as 'R → C', such that, in accepting R, S takes C to be immediately implied and is thereby cognitively compelled to infer C. Importantly, there is nothing in this cognitive warrant that compels S to conclude ˜C on the basis of ˜R. That is to say '˜R → ˜C' need not be a cognitive warrant of S. Indeed, ˜R and C may be entirely consistent not only with each other but with the remainder of S's beliefs also. More importantly, ˜R and C may not seem immediately inconsistent to S and this seems to challenge my interpretation that the defeat of R is an instance of positive undermining.

Why should S's subsequent acceptance of ˜R have any effect whatsoever on her attitude towards C? And what effect should that be?

The answer to the first question is that positive undermining involves forming the positive belief that one's reasons for a belief are no good, not in believing there to be an immediate inconsistency between two claims. The status of R as one of S's reasons for C is not established by S's remembering the inferences she made on the basis of R, but by the significance S takes R to have. That S takes R to be a reason for C is shown by the fact that she took C to be immediately implied by R. There is a cognitive connection in S's mind between R and C. Because of this cognitive connection, the recognition of R's defeat should call to mind the acceptability of C. That C is immediately implied by R (for S) is enough to say that the condition of positive undermining has been met when R is recognized as false by S. Because of this, the undermining of R should have some effect on the cognitive attitude or a reasoner towards C.

Should it require the abandoning of C? No – so long as S has other adequate reasons justifying the belief. (Recall that the normative debate concerning the rationality of belief perseverance in debriefing presupposes that this does not occur and each side agrees on the rationality of the result if it does occur.) What the undermining of a reason should do is call into question the acceptability of the belief thereby altering one's cognitive attitude towards it. In the absence of other, immediately apparent evidence for C, the defeat of R should reduce S's confidence in C. When C is a matter of pressing importance for S (e.g., it is a matter which needs to be settled right away) the defeat of R should also occasion the search for other reasons for C. In some circumstances (e.g., depending on what is at stake) the defeat of R should decrease if not eliminate S's reliance on C, (e.g., when selecting premises for subsequent reasoning). Should S find that all of her reasons for believing C are no good – i.e., that she has no good reason whatsoever for believing that C – then she should be rationally obliged to abandon C altogether. So, while the defeat of R need not require the abandoning of C, it is a case of positive undermining, and to suggest, as the coherence theory does, that no change in S's cognitive attitude is required cannot be taken to be good rational advice.

In summary, the account I have proposed is like Harman's coherence theory in that it does not require that rational agents track their reasons; it "does not suppose there are continuing links of justification dependency that can be consulted when revising one's beliefs."[72] Rather, it claims that part of the significance information has for reasoners is the role it plays in the inferences they make. Further, failure to make the right sorts of inferences with a given piece of information is a failure to understand the significance of that piece of information. The proposed account is consistent with existing psychological explanations of the perseverance effect without sanctioning the behavior is rational. It is perhaps quite understandable that a reasoner might mistake the overall (evidentiary and explanatory) coherence of her beliefs as evidence for the acceptability of some particular belief, C, without realizing that the remainder of her beliefs would cohere equally with C's opposite. Yet, in the absence of a reason to accept C instead of ~C, one's cognitive attitude towards each should not be radically different.

[72] Harman, *Change in View*, 39.

6. Bounded Rationality Revisited

Intuitively it would seem that any principles of rationality that are to guide or regulate our thinking must be principles that we are capable, at least in principle, of following. But this does not mean that our everyday performance has to be sanctioned as rational.

There are several problems with the unqualified claim that principles of rationality must be competence norms. In the first place, how is competence to be determined? Perhaps the day-to-day performance of untutored reasoners is not the best measure of competence. Especially if the habits of mind which serve as the guiding principles of the inferences we make can be altered with training – as is the hope of every course in reasoning skills and critical thinking. The point of teaching and learning reasoning skills is to train the mind to habitually invoke or rely upon good cognitive warrants and to detract from the reliance upon guiding principles of inference that are unsound. Moreover, in judging some of our performances to be rational and others of them not so, we appeal to some standard or ideal which, while we can grasp, we do not always attain. Thus there must be some measure of rationality beyond our own behavior or cognitive habits to which we appeal when conceptualizing rationality.

Second, even if we accept that theorists must presume a background of rational competence against which the performance of individual acts of reasoning are measured, this need not require theorists to presume that all human cognitive habits, strategies and tendencies are rational. It is entirely possible that generally competent human reasoners have a variety of cognitive tendencies which, while wholly or generally unreliable, they nevertheless rely upon with predictable regularity. To suppose that we are rationally competent in general does not mean that there are not systematic ways in which we fail to be rational. Many of these cognitive biases are well-known and have been widely studied. Contrary to Harman's coherence theory,[73] a belief does not acquire justification simply by being believed. Rather, if some of our ways of acquiring or preserving beliefs are not always rational, then the mere fact of believing does on its own count as a reason for the justifiability of the belief, let alone show that belief to be justified. Instead, it must additionally be shown that the believer is being rational in believing what she does. More generally, our descriptions and theories of human reasoning behavior should not exclude the very possibility that some failure of rationality can occur on any particular occasion, even if we accept on principle that it cannot occur in every occasion or even on most occasions.

[73] Harman, *Change in View*, 35.

Finally to challenge the idea that constitutive competencies should serve as the final ground for our normative ideals of rationality, consider the following case. Imagine a group of reasoners who, for whatever reason, are *by our lights* constitutionally incompetent in a certain respect. Perhaps they perennially draw inferences that lead them into self-deception or akrasia despite our best pedagogical efforts and attempts to bring this to their attention. What are we to say of such a group? Should it be said that this group is perfectly rational in their own way, according to their own competence level? Instead, might we not want, while recognizing the cognitive limitations of such a group, to be able to say that they are irrational in certain specifiable ways? Indeed, suppose that there are those among us who *do* track our reasons in certain types of situations, perhaps by taking a mental note of them. Are we then to say that there are two standards of rationality, one for people with good memories and another for those with poor memories? Conceiving of rational norms solely as competence norms, combined with the view that there are different levels of rational competency, leads to the problem of relativism about the norms themselves. To say that our own competencies are the final ground for our rational norms is to say that it is inconceivable that we are somehow constitutionally irrational in certain respects. Yet, as the above example shows, there is no inconsistency in supposing this. Rather, we are rational to the extent that we are capable (constitutionally or otherwise) of acting in accordance with a set of standards and principles which are external to us. If we are unable to live up to those standards because of our cognitive constitutions we may not be faulted for this, but that does not make the behavior rational.

In the end, perhaps there are two morals to this story. First, one way to improve our overall rationality by is lowering our standards and expectations. And second, even proposing competence norms of which we feel ourselves capable involves picturing an ideally competent reasoner, or a set of standards against which our competence can be measured.

SIX SIGNS OF SCIENTISM[1]

Susan HAACK

ABSTRACT: As the English word "scientism" is currently used, it is a trivial verbal truth that scientism – an inappropriately deferential attitude to science – should be avoided. But it is a substantial question when, and why, deference to the sciences is inappropriate or exaggerated. This paper tries to answer that question by articulating "six signs of scientism": the honorific use of "science" and its cognates; using scientific trappings purely decoratively; preoccupation with demarcation; preoccupation with "scientific method"; looking to the sciences for answers beyond their scope; denying the legitimacy or worth of non-scientific (e.g., legal or literary) inquiry, or of writing poetry or making art.

KEYWORDS: scientism, honorific use of "science", demarcation of science, scientific method, science and values

> A man must be downright crazy to deny that science has made many true discoveries. C. S. Peirce[2]
>
> Scientism ... employs the prestige of science for disguise and protection. A.H. Hobbs[3]

Science is a good thing. As Francis Bacon foresaw centuries ago, when what we now call "modern science" was in its infancy, the work of the sciences has brought both light, an ever-growing body of knowledge of the world and how it works, and fruit, the ability to predict and control the world in ways that have both extended and improved our lives. But, as William Harvey complained, Bacon really did write about science "like a Lord Chancellor"[4] – or, as we might say today, "like a promoter," or "like a marketer." Certainly he seems to have been far more keenly aware of virtues of science than of its limitations and potential dangers.

Yet science is by no means a *perfectly* good thing. On the contrary, like all human enterprises, science is ineradicably is fallible and imperfect. At best its progress is ragged, uneven, and unpredictable; moreover, much scientific work is

[1] © 2009 Susan Haack. All rights reserved.

[2] Charles Sanders Peirce, *Collected Papers*, eds. Hartshorne, Charles, Paul Weiss, and (volumes 7 & 8) Arthur Burks (Cambridge: Harvard University Press, 1931-58), 5.172 (1903). References to the *Collected Papers* are by volume and paragraph number.

[3] A. H. Hobbs, *Social Problems and Scientism* (Harrisburg: Stackpole Press, 1953), 17.

[4] My source is Peirce, *Collected Papers* (note 2 above), 5.361 (1877). (Bacon was for a time Lord Chancellor – roughly, what in the U.S. would be called Attorney General – of England.)

unimaginative or banal, some is weak or careless, and some is outright corrupt; and scientific discoveries often have the potential for harm as well as for good – for knowledge *is* power, as Bacon saw, and power can be abused. And, obviously, science is by no means the *only* good thing, nor – only a little less obviously – even the only good form of inquiry. There are many other valuable kinds of human activity besides inquiry – music, dancing, art, story-telling, cookery, gardening, architecture, to mention just a few; and many other valuable kinds of inquiry – historical, legal, literary, philosophical, etc.

As I indicated by giving *Defending Science – Within Reason*[5] its subtitle, *Between Scientism and Cynicism*, we need to avoid *both* under-estimating the value of science, *and* over-estimating it. What I meant by "cynicism" in this context was a kind of jaundiced and uncritically critical attitude to science, an inability to see or an unwillingness to acknowledge its remarkable intellectual achievements, or to recognize the real benefits it has made possible. What I meant by "scientism" was the opposite failure: a kind of over-enthusiastic and uncritically deferential attitude towards science, an inability to see or an unwillingness to acknowledge its fallibility, its limitations, and its potential dangers. One side too hastily dismisses science; the other too hastily defers to it. My present concern, of course, is with the latter failing.

It is worth noting that the English word "scientism" wasn't always, as it is now, pejorative. In the mid-nineteenth century – not long after the older, broader use of the word "science," in which it could refer to any systematized body of knowledge, whatever its subject-matter, had given way to the modern, narrower use in which it refers to physics, chemistry, biology, and so on, but not to jurisprudence, history, theology, and so forth[6] – the word "scientism" was neutral: it meant, simply, "the habit and mode of expression of a man of science." But by the early decades of the twentieth century "scientism" had begun to take on a negative tone – initially, it seems, primarily in response to over-ambitious ideas about how profoundly our understanding of human behavior would be transformed if only we

[5] Susan Haack, *Defending Science – Within Reason: Between Scientism and Cynicism* (Amherst: Prometheus Books, 2003).

[6] According to Friedrich von Hayek, although the earliest example given by Murray's *New English Dictionary* was dated 1867, this narrower usage was already coming into play by 1831, with the formation of the British Association for the Advancement of Science. F. A. Von Hayek, "Scientism and the Study of Society," *Economica* (August 1942): 267, n.2, citing John T. Merz, *History of European Thought in the Nineteenth Century* vol. I (Edinburgh: W. Blackwood and Sons, 1896), 89. See also the entry on "science" in the *Oxford English Dictionary* online (available at http://dictionary.oed.com).

were to apply the methods that had proven so successful in the physical sciences.[7] And by the mid-twentieth century, scientism had come to be seen as a "prejudice,"[8] a "superstition,"[9] an "aberration" of science.[10] By now this negative tone is predominant;[11] in fact, the pejorative connotations of "scientism" are now so thoroughly entrenched that defenders of the autonomy of ethics, or of the legitimacy of religious knowledge, etc., sometimes think it sufficient, instead of actually engaging with their critics' arguments, to dismiss them in a word: "scientistic."

So, as the term "scientism" is usually currently used, and as I shall use it, it is a trivial verbal truth that scientism should be avoided. It is, however, a substantial question exactly *what* it is that is to be avoided – when, and why, deference to the sciences is appropriate and when, and why, it is inappropriate or exaggerated. My primary purpose here is to suggest some ways to recognize when this line has been crossed, when respect for the achievements of the sciences has transmuted into the kind of exaggerated deference characteristic of scientism. These are the "six signs of scientism" to which my title alludes. Briefly and roughly summarized, they are:

> 1. Using the words "science," "scientific," "scientifically," "scientist," etc., honorifically, as generic terms of epistemic praise.
>
> 2. Adopting the manners, the trappings, the technical terminology, etc., of the sciences, irrespective of their real usefulness.
>
> 3. A preoccupation with demarcation, i.e., with drawing a sharp line between genuine science, the real thing, and "pseudo-scientific" imposters.
>
> 4. A corresponding preoccupation with identifying the "scientific method," presumed to explain how the sciences have been so successful.
>
> 5. Looking to the sciences for answers to questions beyond their scope.

[7] See the *Oxford English Dictionary* online (note 6 above) entry on "scientism."

[8] Hayek, "Scientism and the Study of Society" (note 6 above), 269 (describing scientism, the "slavish imitation of the method and language of science" as a "prejudice").

[9] E. H. Hutten, *The Language of Modern Physics* (London: Allen and Unwin, 1956), 273 (describing scientism as "superstitious").

[10] Peter Medawar, "Science and Literature," *Encounter* XXXI.1 (1969): 23 (describing scientism as an "aberration of science").

[11] There are exceptions, such as Michael Shermer, who adopts the word "scientism" as a badge of honor, writing in "The Shamans of Scientism," *Scientific American* 287, 3 (September 2002): 35 that "[s]cientism is a scientific worldview that encompasses natural explanations for all phenomena, eschews supernatural explanations, and embraces empiricism and reason as the twin pillars of a philosophy of life suitable for an Age of Science." But this *is* an exception.

> 6. Denying or denigrating the legitimacy or the worth of other kinds of inquiry besides the scientific, or the value of human activities other than inquiry, such as poetry or art.

I will take these six signs in turn – always trying, however, to keep their interrelations in sight, to signal the mistaken ideas about the sciences on which they depend, and to steer the sometimes very fine line between candidly repudiating scientism, and surreptitiously repudiating science. And, then – taking advantage of the opportunity provided by the last of these signs of scientism – I will comment briefly on some of the tensions between contemporary, scientific culture and the older traditions that, in much of the world, it has by now at least partially displaced.

1. The honorific use of "science" and its cognates

Over the last several centuries, the work of the sciences has enormously enriched and refined our knowledge of the world. And as the prestige of the sciences grew, words like "science," "scientifically," etc., took on an honorific tone: their substantive meaning tended to slip into the background, and their favorable connotation to take center stage. Advertisers routinely boast that "science has shown" the superiority of their product, or that "scientific studies" support their claims. Traditional or unconventional medical treatments are often dismissed out of hand, not as ill-founded or untested, but as "unscientific." Skeptical of some claim, we may ask, not "is there any *good* evidence for that?" but "is there any *scientific* evidence for that?" Needing to craft a test to help judges determine whether expert testimony is reliable enough to be admitted, the U.S. Supreme Court suggests that such testimony must be "scientific knowledge," arrived at by the "scientific method."[12] A historian arguing that there is no foundation in the evidence for the idea that ancient Greek philosophy was borrowed from the Egyptians describes this idea as "unscientific."[13] The titles of conferences and books speak of "Science and Reason,"[14] as if the sciences had a monopoly on reason itself. A recent editorial in

[12] *Daubert v. Merrell Dow Pharms., Inc.*, 509 U.S. 579 (1993). See also Susan Haack, "Trial and Error: The Supreme Court's Philosophy of Science," *American Journal of Public Health* 95 (2005): S66-73; reprinted in Haack, *Putting Philosophy to Work* (Amherst: Prometheus Books, 2008), 161-82.

[13] Mary Lefkowitz, *Not Out of Africa* (New York: Basic Books, 1996), 157.

[14] I am thinking, for example, of the conference at the New York Academy of Sciences in which I participated in 1996, and the corresponding volume. Paul R. Gross, Norman Levitt, and Martin Lewis, eds., *The Flight from Science and Reason* (Baltimore: Johns Hopkins University Press, 1997). I had suggested that the terms be reversed ("Reason and Science") – but my suggestion wasn't taken up.

the *Wall Street Journal* describes studies of charter schools where students are chosen by lottery as "scientific and more reliable" than studies of schools that select their students on merit.[15] The honorific usage is ubiquitous.

Naturally enough, once "science," "scientific," etc., have become honorific terms, practitioners uneasy about the standing of their discipline or approach like to use them emphatically and often. In 1953 Prof. Hobbs provided a splendid list of excerpts from publishers' blurbs for sociology texts: "a scientific approach"; "scientifically faces the problems of … marriage"; "approaches social problems from the … scientific point of view … unassailable [conclusions]"; "sternly scientific"; and so on and on.[16] And nowadays, of course – though departments of physics and chemistry feel no need to stress that what *they* do is science – universities offer classes and degrees in "Management Science,"[17] "Library Science," "Military Science," and even "Mortuary Science."[18]

But this honorific usage of "science" and its cognates leads to all kinds of trouble. It makes it easy to forget that, remarkable as the achievements of the natural sciences have been, not all, and not only, scientists are good, thorough, honest inquirers; it tempts us to dismiss bad science as not really science at all; and it seduces us into the false assumption that whatever is *not* science is no good, or at any rate inferior. Yes, the best scientific work is a remarkable human cognitive achievement; but even this best scientific work is fallible, and there is plenty of good, solid work in non-scientific disciplines such as history, legal scholarship, music theory, etc. – not to mention the vast body of practically useful knowledge accumulated by farmers, sailors, ship-builders, and artisans of every kind, and the considerable resources of knowledge of herbs, etc., embodied in traditional medical practices.[19]

[15] "Do Charters 'Cream' the Best?", *Wall Street Journal*, September 24th (2009): A20.

[16] Hobbs, *Social Problems and Scientism* (note 3 above), 42-3.

[17] For a skeptical view of this supposed discipline, see Matthew Stewart, "The Management Myth," *Atlantic Monthly* 297, 5 (2007): 80-87.

[18] In 1968 C. Trusedell gave a list based on a random search of graduate school catalogues: "'Meat and Animal Science' (Wisconsin), 'Administrative Sciences' (Yale), 'Speech Science' (Purdue), … 'Forest Science' (Harvard), 'Dairy Science' (Illinois), 'Mortuary Science' (Minnesota)." Trusedell, *Essays in the History of Mechanics* (New York: Springer, 1968), 75. The list, and especially "Mortuary Science," became famous among philosophers of science when Jerome Ravetz cited it in *Scientific Knowledge and Its Social Problems* (Oxford: Clarendon Press, 1971), 387, n. 25.

[19] See Dagfinn Føllesdal, "Science, Pseudo-Science and Traditional Knowledge," ALLEA (All European Academies) *Biennial Handbook*, 2002: 27-37; citing Fenstad, E.-J., et al., *Declaration on Science and the Use of Scientific Knowledge*, UNESCO World Conference on Science 2003,

And, inevitably, the honorific use of "science" encourages uncritical credulity about whatever new scientific idea comes down the pike. But the fact is that all the explanatory hypotheses that scientists come up with are, at first, highly speculative, and most are eventually found to be untenable, and abandoned. To be sure, by now there is a vast body of well-warranted scientific theory, some of it *so* well-warranted that it would be astonishing if new evidence were to show it to be mistaken – though even this possibility should never absolutely be ruled out. (Rigid dogmatism is always epistemologically undesirable, rigid dogmatism about even the best-warranted scientific theory included.)[20] But this vast body of well-warranted theory is the surviving remnant of a much, much vaster body of speculative conjectures, most of which came to nothing – a fact which is bound to be obscured if we use "scientific" more or less interchangeably with "reliable, established, solid," and so forth.

2. Inappropriately borrowed scientific trappings

Besides encouraging the honorific use of "science" and its cognates, the successes of the natural sciences have also tempted many to borrow the manners, the trappings, of these fields, in hopes of looking "scientific" – as if technical terminology, numbers, graphs, tables, fancy instruments, etc., were enough by themselves to guarantee success. When Friedrich von Hayek wrote of the "tyranny" that "the methods and technique of the Sciences ... have exercised ... over ... other subjects"[21] he had in mind social scientists' efforts to look as much as possible like physicists – despite their radically different subject-matters. And there certainly *is* something objectionably scientistic about adopting the trappings associated with physics, chemistry, etc., not as useful transferable tools, but as a smoke-screen hiding shallow thinking or half-baked research. Even those who work in disciplines no one would hesitate to classify as sciences sometimes focus too much on form and too little on substance. An epidemiologist testing the side-effects of a morning-sickness drug meticulously calculates the statistical significance of his results, but fails to distinguish women who took the drug during the period of pregnancy when

"Preamble," 4 (available at <http://www.unesco.org/science/wcs/eng/declaration_e.htm>, last visited September 15, 2009).

[20] As I was writing this paper, newly-discovered fossils obliged evolutionary biologists to re-think the ancestry of *homo sapiens* – we are, it now seems, less directly related to chimpanzees than was formerly supposed. See Robert Lee Hotz, "Fossils Shed Light on Human Past," *Wall Street Journal* October 2 (2009): A3.

[21] Friedrich von Hayek, *The Counter-Revolution of Science* (Glencoe: Free Press, 1952), 13.

fetal limbs were forming from those who took it later;[22] another offers impressive-looking tables of cases, but fails to check whether the information in the tables matches the information in the text.[23]

But this kind of misuse of scientific tools and techniques is even commoner in the social sciences, where, as Robert Merton puts it, practitioners only too often "take the achievements of physics as the standard of self-appraisal. They want to compare biceps with their bigger brothers."[24] Lengthy introductory chapters on "methodology" in sociology texts are sometimes only window-dressing; and more often than one would like the graphs, tables, and statistics in social-scientific work focus attention on variables that can be measured at the expense of those that really matter, or represent variables so poorly defined that *no* reasonable conclusion can be drawn. David Abrahamson's Second Law of Criminal Behavior is a classic example: "A criminal act is the sum of a person's criminalistic tendencies plus his total situation, divided by the amount of his resistance," or: "C = (T+ S)/R."[25] The highly mathematical character of contemporary economic theory has contributed to the curious idea that economics is the "Queen of the social sciences" – a title to which psychology[26] would seem to have a much more legitimate claim. But too often those elegant mathematical models turn out to be based on assumptions about "rational economic man" true of no real-world economic actors.[27] And, sadly, policy recommendations based on flawed sociological statistics or flawed economic models often acquire an undeserved status because they are perceived as "science-based."

Inappropriately borrowed scientific trappings are also common in philosophy, where many journals and publishers have adopted such practices as the

[22] Olli P. Heinonen, Denis Slone, and Samuel Shapiro, *Birth Defects and Drugs in Pregnancy* (Littleton: Sciences Group, 1977); see in particular the description of the project design and data collection, 8-29. The record in *Blum v. Merrell Dow Pharms, Inc*, 33 Phila. Co. Rptr., 193 (Ct. Comm. Pleas Pa. 1996), 215-7, shows that Dr. Shapiro admitted under oath that the study had failed to distinguish these two sub-groups of the sample.

[23] Christine Haller and Neal A. Benowitz, "Adverse Cardiovascular and Central Nervous System Events Associated with Dietary Supplements Containing Ephedra Alkaloids," *New England Journal of Medicine* 343 (2000): 1836. (The table is incompatible with the text on the same page.)

[24] Robert Merton, *Social Theory and Social Structure* (1957; enlarged ed., Glencoe: Free Press, 1968), 47.

[25] David Abrahamson, *The Psychology of Crime* (New York: Columbia University Press, 1960), 37.

[26] Of course, psychology also suffers from scientism; and also has a therapeutically-oriented wing in which inquiry takes second place to practice.

[27] See Robert L. Heilbroner, *The Worldly Philosophers* (1958: 7th ed., New York: Simon and Schuster, 1999), chapter xi. Susan Haack, "Science, Economics, 'Vision,'" *Social Research* 71, 2 (2004): 167-83; reprinted in Haack, *Putting Philosophy to Work* (note 12 above), 95-102.

name-date-page-number style of reference used by psychologists, sociologists, etc., and even their preference for the most recent rather than the original dates (often misleading even on its own turf, inherently more so in a discipline where reliance on authority is wholly out of place, and catastrophic when the historical development of an idea matters). Even giving priority to peer-reviewed publication, another practice adopted from the sciences, is a kind of scientism: for peer review is hardly perfect as a rationing device even for scarce space in scientific journals,[28] and is inherently more susceptible to corruption the more a profession is dominated, as philosophy is, by cliques, parties, and schools.[29] And, of course, in philosophy as in the social sciences, technical terminology is far too often not, as it could and should be, a carefully-crafted sign of hard-won intellectual advance, but only self-important jargon designed to attract others to (what you hope will be) a bandwagon.[30]

None of this is to deny, of course, that sometimes scientific tools and techniques turn out also to be genuinely useful to inquirers in other fields: historians use a cyclotron to determine whether the composition of the ink in two earlier printed versions of the bible was the same as that in the "Gutenberg Bible" of 1450-55;[31] they use DNA identification techniques to test the hypothesis that Thomas Jefferson was the father of the children born to his house-slave Sally Hemings;[32] and even borrow medical imaging devices to distinguish the traces of writing on the lead "postcards" on which Roman soldiers wrote home from the marks of centuries of weathering;[33] General Motors uses a model designed by the

[28] See Susan Haack, "Peer Review and Publication: Lessons for Lawyers," *Stetson Law Review* 36 (2007): 789-819.

[29] Nowadays, thinking about the condition of the philosophical journals, I'm afraid I sometimes find this observation of Michael Polanyi's coming unbidden to mind: "if each scientist set out each morning with the intention of doing the best bit of safe charlatanry which would just help him into a good post, there would soon exist no effective standards by which such deception could be detected." Michael Polanyi, *Science, Faith and Society* (Oxford: Oxford University Press, 1946), 40.

[30] See Susan Haack, "The Meaning of Pragmatism: The Ethics of Terminology and the Language of Philosophy Today," *Teorema* XXX/III.3 (2009): 9-29.

[31] It was; and historians now believe that Gutenberg printed all three. See Robert Buderi, "Science: Beaming in on the Past," *Time* Mar. 10 (1986), available at <http://www.time.com/time/printout,0,8816,96050,00.html> (last visited October 1, 2009).

[32] See Jefferson-Hemings Scholars' Commission, *Report on the Jefferson-Hemings Matter* (April 12, 2001); William G. Hyland, Jr., *In Defense of Thomas Jefferson: The Sally Hemings Sex Scandal* (New York: St. Martin's Press, 2009). (The reasonable conclusion seems to be a very modest one: that one of Sally Hemings' children was fathered by some male member of the Jefferson family.)

[33] "Wish You Were Here," *Oxford Today* 10, 3 (1998): 40.

Centers for Disease Control to track an "epidemic" of defects in its cars and trucks.[34] And so on. What is scientistic is not borrowing scientific tools and techniques, as such, but borrowing them, as it were, for display rather than serious use.

3. Preoccupation with "the problem of demarcation"

Once "scientific" has become an honorific term, and when scientific trappings only too often disguise a lack of real rigor, it is almost inevitable that the "problem of demarcation," i.e., of drawing the line between genuine science and pretenders, and with identifying and rooting out "pseudo-science," will loom much larger than it should.

Not surprisingly, as the honorific use of "science" began to take hold in the early decades of the twentieth century, so too did an increasing preoccupation with demarcation: in Logical Positivism (where a key theme was the demarcation of empirically meaningful, scientific work from high-flown but meaningless metaphysical speculation); and, most strikingly, in Karl Popper's philosophy of science.[35] The Positivists had proposed *verifiability* as the mark of the *empirically meaningful*; Popper turned this on its head. Noting that, while no finite number of positive instances could show an unrestricted universal statement true, a single counter-instance is enough to show it false, Popper proposed *falsifiability*, *testability*, or (as he also says) *refutability* as the criterion of demarcation of the genuinely *scientific*.[36] A real scientific theory, according to Popper, can be subjected to the test of experience and, if it is false, can be shown to be false; while a theory that rules nothing out is not a scientific theory at all.

This sounds simple enough. But in fact it never became entirely clear what, exactly, Popper's criterion was, nor what, exactly, it was intended to rule out, nor, most to the present purpose, what exactly – besides the honorific use of "science" – the motivation was for wanting a criterion of demarcation in the first place; in fact, it became increasingly *un*clear. For example, initially it sounded as if Popper intended to exclude Marxist "scientific socialism," along with Freud's and Adler's psycho-analytic theories, as unfalsifiable. But in *The Open Society and Its Enemies* (1945) Popper grants that, after all, Marxism *is* falsifiable – in fact, it was falsified by

[34] Gregory L. White, "GM Takes Advice from Disease Sleuths to Debug Cars," *Wall Street Journal*, 8 April (1999): B1, B4.

[35] The origins of this idea are described in Karl R. Popper, *Unended Quest* (La Salle: Open Court, 1979), 31-38 (published as a book after first appearing in *The Philosophy of Karl Popper*, ed. Paul A. Schilpp (La Salle, IL: 1974), 3-181.

[36] Karl R. Popper, *The Logic of Scientific Discovery* (1934; English ed., London: Routledge, 1959).

the events of the Russian revolution.[37] What went wrong was not that the theory was unfalsifiable but that, instead of abandoning their theory in the face of contrary evidence, Marxists made *ad hoc* modifications to save it. So Popper's supposedly logical criterion was transformed into a partly methodological test – a test, moreover, according to which badly conducted science is not really science at all.

Again: for a long time Popper claimed that his criterion of demarcation excluded the theory of evolution; which, he wrote, is not a genuine scientific theory but a "metaphysical research programme."[38] Then he changed his mind: evolution *is* science, after all.[39] And again – quietly shifting from writing of falsifiability as a criterion of the scientific to suggesting that it is a criterion of the empirical – Popper acknowledged that the category of "non-science" includes not only pseudo-science, but also such legitimate but non-empirical areas of inquiry as metaphysics and mathematics.[40] By the time you notice that he describes his criterion as a "convention,"[41] and even, in the introduction to the English edition of *The Logic of Scientific Discovery*, writes that scientific knowledge is continuous with everyday empirical knowledge,[42] you can hardly avoid the conclusion that the apparently simple idea he started with has become something of an intellectual monster.

With the benefit of hindsight, it looks as if Popper's criterion of demarcation proved so attractive to so many in part because it was amorphous – or rather, polymorphous – enough to seem to serve a whole variety of agendas: such as federal courts' interest in distinguishing reliable scientific testimony from "junk science,"[43] or in determining whether "creation science" is really science, and hence may constitutionally be taught in public high schools.[44] Other criteria have been

[37] Karl R. Popper, *The Open Society and Its Enemies* (1945; revised ed., 1950), 374.

[38] Popper, *Unended Quest* (note 34 above), 167-180.

[39] Karl R. Popper, "Natural Selection and Its Scientific Status," a lecture of 1977, first published in *Dialectica* 32 (1978); reprinted in *A Pocket Popper*, ed. David Miller (London: Fontana, 1983), 239-246.

[40] Popper, *The Logic of Scientific Discovery* (note 35 above), 39.

[41] *Ibid.*, 37.

[42] *Ibid*, 18.

[43] *Daubert* (1993) (note 12 above). Of course, though the Supreme Court doesn't realize this, it is hard to think of a philosophy of science less suitable than Popper's – which expressly denies that any scientific theory is ever shown to be reliable – to serve as a criterion of reliability. See Susan Haack, "Federal Philosophy of Science: A Deconstruction – and a Reconstruction," *NYU Journal of Law & Liberty*, 5.2 (2010): 394-435.

[44] *McLean v. Arkansas Board of Education*, 529 F.Supp.1255 (1982). Of course, though the court in *McLean* didn't realize this, in view of Popper's ambivalence about the status of the theory of

proposed – that real science relies on controlled experiments, for example (which, however, would rule out not only anthropology and sociology, but also – most implausibly of all – astronomy). The best we might hope for, I believe, is a list of "signs of scientificity" none of which would be shared by all sciences, but each of which would be found, in some degree, in some sciences. The fact is that the term "science" simply *has no* very clear boundaries: the reference of the term is fuzzy, indeterminate and, not least, frequently contested.

This is not to say that we can't, in a rough and ready way, distinguish between the sciences and other human activities, including other human cognitive activities; but it is to say that any such distinction can *only* be rough and ready. I might say, as a first approximation, that science is best understood, not as a body of knowledge, but as a kind of inquiry (so that cooking dinner, dancing, or writing a novel, isn't science, nor pleading a case in court). At a second approximation, I would add that, since the word "science" has come to be tied to inquiry into empirical subject-matter, formal disciplines like logic or pure mathematics don't qualify as sciences, nor normative disciplines like jurisprudence or ethics or aesthetics or epistemology). And at a third approximation, to acknowledge that the work picked out by the word "science" is far from uniform or monolithic, it makes sense to say, rather, that the disciplines we call "the science*s*" are best thought of as forming a loose federation of interrelated kinds of inquiry.

But if we want to get a clear view of the place of the sciences among the many kinds of inquiry, of the place of inquiry among the many kinds of human activity, and of the interrelations among the various disciplines classified by deans and librarians as sciences, we will need to look for continuities as well as differences. For there are marked affinities between (as we say) "historical" sciences like cosmology and evolutionary biology, and what we would ordinarily classify simply as historical inquiry. There is no sharp boundary between psychology and philosophy of mind, nor between cosmology and metaphysics.[45] Nor is there any very clear line between the very considerable body of knowledge that has grown out of such primal human activities as hunting, herding, farming, fishing, building, cooking, healing, midwifery, child-rearing, etc., etc., and the more systematic knowledge of agronomists, child psychologists, etc.[46]

evolution it is far from clear that his criterion would enable us to classify evolution as science, and creation "science" as non-science.

[45] See Susan Haack, "Not Cynicism but Synechism: Lessons from Classical Pragmatism" (2005), in Haack, *Putting Philosophy to Work* (note 12 above), 79-93.

[46] For that matter, there are also some very significant differences among the various disciplines conventionally classified as sciences – between the natural and the social sciences, of course, but also between physics and biology, between sociology and economics, and so on.

Scientific inquiry is recognizably continuous with more commonplace and less systematic kinds of empirical inquiry – inquiry into the causes of spoiled crops, the design of fishing boats, the medicinal properties of herbs, etc.. It is more systematic, more refined, and more persistent; but sometimes it rediscovers, and builds on, traditional knowledge: as Linnaeus, for example, built on traditional Lap taxonomies of plants and animals;[47] or as many drugs now part of the arsenal of modern scientific medicine were derived from what were originally folk remedies. An example would be digitalis, extracted from a plant called the foxglove: long used as a folk remedy, digitalis was first named in 1542; its clinical properties were first described by William Withering in 1785; and by the mid-twentieth century it was in common use by physicians for the treatment of heart ailments.[48]

Suppressing the demarcationist impulse enables us to see the Popperian requirement that a theory rule something out, that it not be compatible with absolutely *anything* and *everything* that might happen, for what it really is: a mark, not of its being scientific specifically, but of its being genuinely explanatory. And willingness to take contrary evidence seriously can also be seen for what it really is: a mark not, as Popper supposes, of the scientist specifically, but of the honest inquirer, in whatever field. (The historian who ignores or destroys a document that threatens to undermine his favored hypothesis is guilty of just the same kind of intellectual dishonesty as the scientist who ignores or fails to record the results of an experiment that threatens to falsify his theory.) "Scientism," as Hayek shrewdly observes, confuses "the general spirit of disinterested inquiry" with the methods and language of the natural sciences.[49]

And suppressing the demarcationist impulse will also have the healthy effect of obliging us to recognize poorly-conducted science as just that, poorly-conducted *science*; and of encouraging us, instead of simply sneering at "pseudo-science," to

[47] I learned this from Føllesdal, "Science, Pseudoscience and Traditional Knowledge" (note 19 above); Føllesdal again cites the 2002 UNESCO report (note 19 above).

[48] Jeremy N. Norman, "William Withering and the Purple Foxglove: A Bicentennial Tribute," *Journal of Clinical Pharmacology* 25 (1985): 479-83. Susan Wray, D. A. Eisner, and D. G. Allen, "Two Hundred Years of the Foxglove," *Medical History*, *Supplement* 5 (1985): 132-50. Dale Groom, "Drugs for Cardiac Patients," *American Journal of Nursing* 56, 9 (September 1956): 1125-1127. James E. F. Reynolds, ed., *Martindale: The Extra Pharmacopoeia* (London: Pharmaceutical Press, 30th ed., 1993), 665-6. Another example would be quinine, derived from the bark of the cinchona tree, now standard in the treatment of malaria. See Tropical Plant Database file for quinine (available at <http://www.rain-tree.com/quinine.htm>, last visited October 6, 2009); Lexi Krock, "Accidental Discoveries" (available at <http://www.pbs.org/wgbh/nova/cancer/discoveries.html>, last visited October 6, 2009).

[49] Hayek, *The Counter-Revolution of Science* (note 21 above), 15.

specify what, exactly, is wrong with the work we are criticizing: perhaps that it is too vague to be genuinely explanatory; perhaps that, though it uses mathematical symbolism or graphs or fancy instruments, these are purely decorative, and do no real work; perhaps that claims which are thus far purely speculative are being made as confidently as if they were well-warranted by evidence; and so on. If we still had a use for the term "pseudo-science," it might be best reserved to refer to such public-relations exercises as the Creation Science "movement" – what a revealing word! – which, so far as I can tell, really involves no real inquiry of any kind.

4. The quest for "scientific method"

The preoccupation with demarcation in turn encourages (and is encouraged by) the idea that real scientific inquiry, the genuine article, differs from inquiry of other kinds in virtue of its uniquely effective method or procedure – the supposed "scientific method." However, we have yet to see anything like agreement about what, exactly, this supposed method *is*. A whole range of different, and incompatible, candidates have been proposed: various forms of inductivism (from an older, stronger version according to which scientists arrive at their hypotheses by induction from observed instances, to more recent, weaker versions according to which scientists arrive at hypotheses by a process better described as imaginative than as inferential, but then test them inductively); various forms of deductivism (Popper's conception of scientific method as a matter of "conjecture and refutation," i.e., making an informed guess, deducing its consequences, and then trying to falsify it, and Imre Lakatos's quasi-Popperian, post-Kuhnian distinction of regressive versus progressive research programs); and, most recently, Bayesian, decision-theoretic, etc., approaches.

Already by 1970 Paul Feyerabend famously drew the radical conclusion that the only methodological principle that would not impede the progress of science is "anything goes."[50] Other philosophers of science have suggested, somewhat more plausibly, that there is no *constant* scientific method, only a method that shifts and changes as science proceeds; or that there is no *single* scientific method, but many different scientific method*s* in different areas of science. But a thoughtful physicist had put his finger on the essential point as early as 1949. "There is a good deal of ballyhoo about scientific method," Percy Bridgman wrote; though, as he shrewdly observed, "the people who talk most about it are the people who do least about it." But no working scientist, he continued, ever asks himself whether he is being "scientific" or using the "scientific method." No: "he is too much concerned with

[50] Paul K. Feyerabend, *Against Method* (London: New Left Books, 1970).

getting down to brass tacks to be willing to spend his time on generalities."[51] "[I]nsofar as it is a method," Bridgman comments, the scientific method is a matter simply of "doing one's damnedest with one's mind, no holds barred."[52]

These bracingly commonsense observations are exactly right. *Any* serious empirical inquirer, whatever his subject-matter, will make an informed guess at the possible explanation of the event or phenomenon that puzzles him, figure out the consequences of that guess, see how well those consequences stand up to the evidence he has and any further evidence he can lay his hands on, and then use his judgment whether to stick with the initial guess, modify it, drop it and start again, or just wait until he can figure out what further evidence might clarify the situation, and how to get it. Over centuries of work, however, scientists have gradually developed an array of tools and techniques to amplify and refine human cognitive powers and overcome human cognitive limitations: techniques of extraction, purification, etc.; instruments of observation from the microscope and the telescope to the questionnaire; mathematical techniques from the calculus to statistics to the computer; and even internal social arrangements that – up to a point, though only up to a point – provide incentives for good, imaginative, honest work, and disincentives to sloppiness and cheating.[53]

The underlying procedures of all serious empirical inquiry – taking a stab at an answer, and then checking it out[54] – are not used *only* by scientists; the scientific "helps" to inquiry, which are constantly being adapted and improved, and are often local to some specific area of science, are not used by *all* scientists. So there *is no* "scientific method" used by all and only scientists. But, far from suggesting that it is simply a mystery how the natural sciences can have "made many true discoveries," this approach suggests a plausible account of how they have gradually managed to refine, amplify, and extend unaided human cognitive powers. It also throws some light on whether the social sciences really use the same method as the natural sciences, or a distinctive method of their own. Like natural-scientific inquiry, social-scientific inquiry will follow the underlying pattern of all serious empirical inquiry. Like natural-scientific inquiry, it will benefit from internal social arrangements than encourage good, honest, thorough work, and discourage

[51] Percy Bridgman, "On Scientific Method" (1949), in Bridgman, *Reflections of a Physicist* (New York: Philosophical Library, 1955), 81.

[52] Percy Bridgman, "The Prospect for Intelligence"(1945), in Bridgman, *Reflections of a Physicist* (note 50 above), 535.

[53] These ideas are developed in detail in Haack, *Defending Science – Within Reason* (note 5 above), chapter 4.

[54] Calling this underlying pattern the "hypothetico-deductive method," as if it were a special, technical procedure, and peculiar to science, is itself a kind of scientism.

cheating. But at least many of the special tools and techniques of which it will have need are likely to be very different from the special tools and techniques most useful in the natural sciences.[55]

5. Looking to the sciences for answers to questions beyond their scope

There are many questions clearly within the scope of one or another of the disciplines conventionally classified as sciences to which there are as yet no warranted answers. (This is why credulity about current scientific speculation, even flimsy and as yet untested speculation, is itself a sign of scientism.) There are also many questions within the scope of the sciences which it is not yet possible even to ask – as once, before DNA was identified and the concept of macromolecule worked out,[56] questions about the structure and function of DNA the answers to which are now known were not so much as conceivable. Still, all these are questions clearly within the scope of the disciplines conventionally classified as sciences; and looking to the relevant sciences to answer them is entirely proper. But there are also are many legitimate questions outside the scope of the sciences altogether: legal, literary, culinary, historical, political, etc., questions – and philosophical questions, on which I shall focus here.

Some issues once within the purview of philosophy of mind or the epistemology of perception have proven susceptible to treatment by the science of psychology; the baffling metaphysical question, "why is there something rather than nothing?" has in part been resolved as cosmologists have tackled the problem of (what they call) "the accretion of matter."[57] Such boundary-shifting is not always or necessarily scientistic – indeed, it has often been a real intellectual advance. But when scientific answers that leave central elements of the older questions untouched are taken to be sufficient, this is scientism.

Results from the sciences frequently have a bearing on questions of policy: environmental science might tell us what the consequences of damming this river would be, medical science at what stage a human fetus becomes viable, social-scientific studies the consequences of changing tax incentives in this way or that, of

[55] These ideas are developed in detail in Haack, *Defending Science – Within Reason* (note 5 above), chapter 6.

[56] The stuff we now call "DNA" was discovered in 1859 by Friedrich Miescher (who called it "nuclein"). The concept of a macromolecule was introduced by Hermann Staudinger in 1922. See Franklin H. Portugal and Jack S. Cohen, *A Century of DNA: A History of the Discovery of the Structure and Function of the Genetic Substance* (Cambridge: MIT Press, 1977); Robert Olby, *The Path to the Double Helix* (Seattle: University of Washington Press, 1974).

[57] See John Maddox, *What Remains to be Discovered: Mapping the Secrets of the Universe, the Origins of Life, and the Future of the Human Race* (New York: Simon and Schuster, 1998), 25 ff.

increasing the number of charter schools, of abolishing the death penalty, or, etc. But though a good deal of scientific work is policy-relevant, scientific inquiry – if it is to be genuine inquiry, and not what is oxymoronically called "advocacy research" – is policy-neutral. Environmental science can't, by itself, tell us whether the benefits of damming the river outweigh the drawbacks, and certainly not whether building the damm is a good idea; medical science can't, by itself, tell us whether abortion is morally acceptable (nor, of course, whether it should be legally permitted); economics can't, by itself, tell us whether we should change the tax system in this way or that. To be sure, environmental scientists, sociologists, economists, etc., will probably have opinions about the policy questions on which their scientific work has a bearing; and it is entirely legitimate for them to express such opinions publicly. But something goes wrong when they allow their ethical or political convictions to affect their judgment of the evidence, or when they present those ethical or political convictions as if they were scientific results.

These relatively simple arguments suggest a relatively simple conclusion: that results from the sciences can give us information about the relation of means to ends, but cannot by themselves tell us what ends are desirable. This is true, so far as it goes; but it doesn't go nearly far enough. It leaves a much deeper matter – whether, and if so, how, scientific results can have *any* bearing on questions about what ends are desirable – untouched. And on this deeper matter, I'm with John Dewey, who wrote that "restoring integration ... between a man's beliefs abut the world in which he lives and his beliefs about the values and purposes that should direct his conduct is the deepest problem of modern life"[58]: the idea of science as *purely* factual, as *entirely* "value-free," and an *wholly irrelevant* to normative questions, is far too crude.

Here (setting aside questions about epistemological, aesthetic, etc., values), I will focus on the ethical. As I see it, ethics is neither a wholly autonomous, a priori discipline, nor simply as a sub-branch of the human sciences. (This is a kind of modest ethical naturalism, informed by the idea that what is good or right for humans to do cannot be entirely divorced from what is good for humans.) Knowledge of what truly enables human flourishing – knowledge to which not only biology but also psychology, sociology, economics, etc., might contribute – though never *sufficient by itself* to tell us what to do, can have *contributory relevance* to ethical questions.

A recent paper in the *Lancet* provides a vivid illustration of the pitfalls of appealing to scientific results as if they were sufficient to answer ethical questions. The authors' thesis is that the morally best system for allocating scarce medical

[58] John Dewey, *The Quest for Certainty* (1929; reprinted, New York: Capricorn Books, 1960), 255.

resources is the "complete lives" principle, which gives priority to adolescents and young adults over infants and older adults. As evidence, they cite empirical surveys showing that "most people think" that the death of an adolescent is worse than the death of an infant.[59] Set aside the fact that they cite only two such studies, neither of which actually reports exactly what their summary suggests.[60] The essential point is that "*most people think* x is morally best" and "x *is* morally best" are different propositions altogether.[61] Conflating them is a sure sign of scientism.

The "evolutionary ethics" offered by E. O. Wilson looks at first blush like another example, albeit a more sophisticated example, of the same kind of scientism. The definition of the moral sentiments, Wilson tells us, falls to experimental psychology, investigation of the heritability of these sentiments to genetics, investigation of the development of the moral sentiments to anthropology and psychology,[62] and the "deep history of the moral sentiments" to evolutionary biology.[63] If the claim is that such scientific investigations are *all* that ethical theory requires, it is surely mistaken: it rests on the unargued presumption that ethics must be understood in terms of moral sentiments; it doesn't tell us *what* sentiments are moral; and, in and of itself the fact (supposing it *is* a fact) that these sentiments can be given an evolutionary explanation does not by itself show that they are, or that they aren't, ethically desirable. It is a kind of scientism.

[59] Govind Persad, Alan Wertheimer, and Ezekiel J. Emanuel, "Principles for Allocation of Scarce Medical Resources," *The Lancet* 373, Jan 31 (2009): 423-31. (Mr. Emanuel is health adviser to President Obama.)

[60] Aki Tsuchiya, Paul Dolan, and Rebecca Shaw, "Measuring People's Preferences Regarding Ageism in Health: Some Methodological Issues and Some Fresh Evidence," *Social Science and Medicine* 57 (2007):688-96 (finding that people are broadly in favor of giving priority to older over younger patients, but noting that how the questions are asked may affect the upshot); Jeff Richardson, "Age Weighting and Discounting: What Are the Ethical Issues?", Working Paper 108, Health Economics Unit, Monash University (Australia) (using the term "empirical ethics" to refer to surveys of people's *beliefs about* ethical questions).

[61] The authors of the *Lancet* paper also fudge the relation of economic values to ethical ones. Perhaps there is a plausible economic argument that society has made a greater economic investment in adolescents or young adults than in infants, and can expect greater future return on the investment in adolescents or young adults than on older people; but Persad et al. simply dismiss the economic fact that society has invested less in underprivileged adolescents or young persons – this irrelevant, they say, because it is itself the result of "social injustice." "Measuring People's Preferences" (note 58 above), 428.

[62] While I was writing this paper a new book suggested fascinating conjectures about the origins of empathy, in humans and other animals. Frans de Waal, *The Age of Empathy* (New York: Harmony, 2009). See also Robert Lee Hotz, "Tracing the Origins of Human Empathy," *Wall Street Journal*, September 26 (2009): A11.

[63] E. O. Wilson, *Consilience: The Unity of Knowledge* (1998; reprinted New York:Vantge, 1999), 279.

But Wilson's evolutionary ethics is one aspect of a larger picture of what he calls "the unity of knowledge"; and his understanding of this "unity" is ambiguous in a crucial way. Sometimes he seems to be offering only the modest thesis that all knowledge must ultimately fit together in a coherent whole (which is obviously true); at other times, the much more ambitious thesis that all knowledge must ultimately be derivable from scientific knowledge (which is – I believe, no less obviously – false). So perhaps it is not entirely surprising that, after seeming to suggest that results from the biological sciences might be sufficient to answer ethical questions, Wilson goes on to ask how the moral instincts can be ranked and which are best subdued, which moral principles are best incorporated into law and which admit of exceptions, etc.[64] This is as much as to acknowledge that biology is relevant but not, after all, sufficient; which, by my lights, is *not* inappropriate, and *not* scientistic, but potentially a step in the right direction.

6. Denigrating the Non-Scientific

Steven Weinberg writes of the gradual "demystification" of the world through scientific advances.[65] And indeed, developments in cosmology and evolutionary biology have provided natural explanations of phenomena once thought to demand *super*natural explanations; and in the process, have shown that questions about "design," whether of organs such as the eye, or of the universe generally, rest on false presuppositions. To acknowledge this is not, by my lights, scientistic. But it *is* scientistic to imagine that advances in the sciences will eventually displace the need for any other kind of inquiry.

Here as elsewhere, the line between appropriate respect for science and inappropriate deference is often a fine one. It is not scientistic to value well-conducted empirical studies of the effects of legal changes (e.g., of the effect of abolishing the death penalty on the murder rate, or the effects of imposing a cap on punitive damages in medical-malpractice suits on the number of physicians a state attracts). It is, however, scientistic to assume that social-scientific "empirical legal studies" are inherently more valuable than traditional interpretive legal scholarship. Again, it is not necessarily objectionable for a university to give priority to medical research with the potential to improve health significantly over other, less practical, research; but it is a real loss – and not only because it is so unpredictable what work really will have important practical applications – if universities cease to value serious intellectual work for its own sake, regardless of subject-matter or potential payoff.

[64] Wilson, *Consilience*, 279-80.

[65] Steven Weinberg, *Dreams of a Final Theory* (1992; reprinted New York, Vintage, 1993), 245.

Moreover, though our capacity for inquiry is a remarkable human talent – strikingly manifested in the sciences, though not only in the sciences – we humans have other talents, too: for story-telling, for singing, dancing, painting, ..., and so on. (It has been conjectured, in fact. that the human capacity for speech – without which neither science nor story-telling would be possible – may have arisen out of a more primitive musical capacity.)[66] Focusing for the moment on story-telling, I note that, loose talk of "two cultures" notwithstanding,[67] there are significant similarities as well as significant differences between science and literature. As Peirce observes, there is nothing more necessary to scientific work than imagination – though the scientific man, he continues, "dreams of explanations and laws,"[68] while a novelist dreams of imaginary people, events, and worlds. By my lights, not only is it scientism to assume that scientific inquiry is inherently better than other kinds of inquiry; it is also scientism to assume that science is inherently more valuable than literature (or art, or music, or, etc.). "Which is more important, science or literature?" is a hopelessly misguided question – as hopelessly misguided as "Which is more important, a sense of humor or a sense of justice?"

*

* *

What we now call "modern science" arose in Europe, and was the work mostly of white men. Post-colonialist, feminist, and other "science critics" sometimes complain that science is racist and sexist – a white male thing. This is a silly idea. Modern science grows out of much older human efforts to understand the world; there were many important anticipations of modern science: in China, in the Arab world, and elsewhere; and by now there are capable scientists of virtually every race and gender. Science isn't a white male thing; it's a *human* thing – as I was forcibly reminded, not long ago, when I talked at length with two post-docs working at a medical research institute in Switzerland,[69] a young woman from Canada, and a young man from Uzbekistan: culturally worlds apart, they shared a common scientific heritage, and common scientific aspirations.

But of course, modern science is also a (relatively) *recent* thing. Moreover, scientific advances can pose a real threat to comfortable ideas about ourselves and

[66] Robert Lee Hotz, "Magic Flute: Primal Find Sings of Music's Mystery," *Wall Street Journal*, July3-5 (2009): A9.

[67] C. P. Snow, "The Two Cultures" (1959), in *The Two Cultures and a Second Look* (Cambridge: Cambridge University Press, 1964).

[68] Peirce, *Collected Papers* (note 2 above), 1.48 (c.1896).

[69] The Friedrich Miescher Institute, Basel. (Recall from note 55 that it was Miescher, a native of Basel, who discovered DNA.)

our place in the universe, and to familiar, traditional ways of doing things. So it should come as no surprise that such advances sometimes meet resistance from those who value older ways. Sometimes, the resistance is foolish. I read, for example, that some prominent Indian social scientists favor the traditional custom of variolation – inoculation with human smallpox matter, accompanied by prayers to the goddess smallpox – over the modern scientific practice of vaccination using cowpox vaccine, which is much less likely to cause smallpox in the patient.[70] This, in my view, is worse than silly.

Nonetheless, it must be frankly acknowledged that when older traditions are displaced by newer, scientific practices and methods, there can be loss as well as gain. (I say "newer, scientific practices and methods"; but I am uncomfortably aware that discriminating the effects of scientific advance from the effects of industrialization, of urbanization, and now of globalization, is formidably difficult, and perhaps not even possible.) Once, the Panare Indians of Venezuela worked together to clear trees with stone axes; with the introduction of new, labor-saving steel axes they could clear trees much faster and more efficiently – but the traditional, agreeably cooperative ways of working died out.[71] Affluent American consumers who appreciate the solidity and craftsmanship of their old-fashioned, low-tech construction techniques sometimes seek out Amish builders to work for them.[72] Academics notice with dismay that students with the vast resources of the internet available to them seem to have forgotten, if they ever knew, how to read an actual book. Virtually all of us, probably, have benefited in one way or another from advances in medical science; many of us, I suspect, like myself, also feel some unease about the impersonal character of technologically sophisticated modern medicine.

Such examples could be multiplied almost without limit; but I will stop here, with a simple thought: that to forget that the technological advances that science

[70] See Meera Nanda, "The Epistemic Charity of Social Constructivist Critics of Science and Why the Third World Should Reject the Offer," in *A House Built on Sand: Exposing Post-Modern Myths about Science*, ed. Noretta Keortge (New York: Oxford University Press, 1998), 291; Nanda cites Fredérique Apfel Marglin, "Smallpox in Two Systems of Knowledge," in *Dominating Knowledge: Development, Culture and Resistance*, eds. Fredérique Apfel Marlin and Stephen Marglin (Oxford: Clarendon Press, 1990), 102-44.

[71] Katherine Milton, "Civilization and Its Discontents: Amazonian Indians," *Natural History* 101, 3, March (1992): 36-42.

[72] "Amish" refers to a religious sect that eschews modern technology, still using horses and buggies instead of motor vehicles, etc. Nancy Keates, "From Barn Raisings to Home Building: Consumers Hire Amish Builders, Citing Craftsmanship, Costs," *Wall Street Journal*, August 15 (2008): W1.

brings in its wake, much as they have improved our lives, have also sometimes come at a real cost in the displacement of valuable traditional practices and skills, is itself a kind of scientism.[73]

[73] My thanks to Mark Migotti for very helpful comments on a draft, and to Pamela Lucken for help in finding relevant material.

AN ARROVIAN IMPOSSIBILITY THEOREM FOR THE EPISTEMOLOGY OF DISAGREEMENT

Nicholaos JONES

ABSTRACT: According to conciliatory views about the epistemology of disagreement, when epistemic peers have conflicting doxastic attitudes toward a proposition and fully disclose to one another the reasons for their attitudes toward that proposition (and neither has independent reason to believe the other to be mistaken), each peer should always change his attitude toward that proposition to one that is closer to the attitudes of those peers with which there is disagreement. According to pure higher-order evidence views, higher-order evidence for a proposition always suffices to determine the proper rational response to disagreement about that proposition within a group of epistemic peers. Using an analogue of Arrow's Impossibility Theorem, I shall argue that no conciliatory and pure higher-order evidence view about the epistemology of disagreement can provide a true and general answer to the question of what disagreeing epistemic peers should do after fully disclosing to each other the (first-order) reasons for their conflicting doxastic attitudes.

KEYWORDS: Arrow's theorem, epistemology, higher-order evidence, peer disagreement, rationality

1. Views about the Epistemology of Disagreement

Situations often arise in which we find ourselves disagreeing with our peers. Even when we have access to the same evidence and respond to that evidence in equally reliable ways, we sometimes form conflicting beliefs. This occurs, for example, when jurors reach different judgments about a defendant's guilt; when meteorologists offer competing weather forecasts; when philosophers do metaphysics; when scientists offer conflicting accounts of experimental data; when physicians pronounce different causes for the same diseases; when politicians make different policy recommendations for addressing social issues; and so on.[1]

[1] Richard Feldman, "Epistemological Puzzles about Disagreement," in *Epistemology Futures*, ed. Stephen Hetherington (Oxford: Oxford University Press, 2006), 216-236; Thomas Kelly, "Peer Disagreement and Higher-Order Evidence," in *Disagreement*, eds. Richard Feldman and Ted Warfield (Oxford: Oxford University Press, 2010), 111-174.

There are several competing views about the proper rational response to disagreement within a group of epistemic peers. According to *conciliatory* views, when epistemic peers have conflicting doxastic attitudes toward a proposition and fully disclose to one another the reasons for their attitudes toward that proposition (and neither has independent reason to believe the other to be mistaken), each peer should always change his attitude toward that proposition to one that is closer to the attitudes of those peers with which there is disagreement.[2] *Steadfast* views, in contrast, maintain that when epistemic peers have conflicting doxastic attitudes toward a proposition and fully disclose to one another the reasons for their attitudes toward that proposition, sometimes some peers may maintain their original attitude toward the proposition.[3]

There are also competing views about the kind of evidence that determines the proper rational response to disagreement within a group of epistemic peers. *First-order evidence* for a proposition is any evidence that bears directly on that proposition's truth-value. First-order evidence can include perceptual evidence, testimonial evidence, inferential evidence, intuition, and so on. *Higher-order evidence* for a proposition, in contrast, is evidence about first-order evidence for that proposition.[4] For example, a person's higher-order evidence for the proposition that God exists might include the fact that a peer takes the ontological argument to be sound, the fact that another peer takes the evidential problem of evil to conclusively refute God's existence, the fact that a peer takes reports of personal experience to be evidence for God's existence, and so on. Higher-order evidence is evidence about what first-order evidence supports.

There is disagreement about whether higher-order evidence for a proposition always suffices to determine the proper rational response to disagreement about that proposition within a group of epistemic peers. According to what I shall call *pure higher-order evidence* (HOE) views, it does. For example, according to the equal weight view, when two peers adopt conflicting doxastic attitudes toward a proposition after full disclosure, the rational response to that disagreement depends upon what those attitudes are and nothing more. *Mixed evidence* views, in contrast, maintain that sometimes first-order evidence about a proposition helps to

[2] See David Christensen, "Epistemology of Disagreement: The Good News," *Philosophical Review* 116 (2007): 187-217; Adam Elga, "Reflection and Disagreement," *Noûs* 41 (2007): 478-502.

[3] I take this terminology from David Christensen, "Disagreement as Evidence: The Epistemology of Controversy," *Philosophy Compass* 4/5 (2009): 756. See also Kelly, "Peer Disagreement and Higher-Order Evidence" and Thomas Kelly, "The Epistemic Significance of Disagreement," in *Oxford Studies in Epistemology*, Volume I, eds. Tamar Szabò Gendler and John Hawthorne (Oxford: Oxford University Press, 2005), 167-196.

[4] See Kelly, "The Epistemic Significance of Disagreement," 185-190.

determine the proper rational response to disagreement. Examples of what I am calling mixed views are Kelly's total evidence view and Lackey's justificationist view.[5]

I shall argue that no view that is both conciliatory and pure HOE can provide a true and general answer to the question of what disagreeing epistemic peers should do after fully disclosing to each other the (first-order) reasons for their conflicting doxastic attitudes. As a matter of principle, any such view is committed to two constraints about the way in which the rational response to disagreement among epistemic peers is a function of those peers' higher-order evidence. These constraints, and an additional adequacy condition for all views of peer disagreement, are formal analogues to the ones that appear in Arrow's Impossibility Theorem.[6] This analogy, together with replies to potential objections, show that conciliatory pure HOE views are either false or unacceptably ad hoc.

I begin, in the first section, with some preliminaries about how to understand the formal structure of peer disagreement situations in a way that makes Arrow's Theorem relevant to the epistemological debate. Next, I motivate an adequacy condition for views about peer disagreement. Then I argue that conciliatory pure HOE views are committed to two additional constraints about the way in which the rational response to disagreement among epistemic peers is a function of those peers' higher-order evidence. After presenting a formal analogue of Arrow's Impossibility Theorem, I consider some ways in which conciliatory pure HOE views might attempt to avoid the upshot of Arrow's Impossibility Theorem as applied to the epistemology of disagreement. I conclude that such views are false if as they cannot avoid the theorem, and unacceptably ad hoc if they can.

2. Abstract Structure of Peer Disagreement Situations

There are at least three doxastic attitudes possible toward any proposition. These attitudes might be course-grained: believing the proposition; disbelieving it (believing it is false); and withholding judgment about it (neither believing nor disbelieving it).[7] They might be fine-grained, such as attitudes that involve

[5] In Kelly, "Peer Disagreement and Higher-Order Evidence" and Jennifer Lackey, "What Should We Do When We Disagree?" in *Oxford Studies in Epistemology*, Volume 3, eds. Tamar Szabò Gendler and John Hawthorne (Oxford: Oxford University Press, 2010), 274-293.

[6] See Kenneth J. Arrow, "A Difficulty in the Concept of Social Welfare," *The Journal of Political Economy* 58 (1950): 328-346; David Austen-Smith and Jeffrey S. Banks, *Positive Political Theory I: Collective Preference* (University of Michigan Press, 1999); John Geanakoplos, "Three Brief Proofs of Arrow's Impossibility Theorem," *Economic Theory* 26 (2005): 211-215.

[7] See Jane Friedman, "Suspended Judgment," *Philosophical Studies* (forthcoming). DOI: 10.1007/s11098-011-9753-y.

confidence levels: believing the proposition with degree of confidence 1; believing it with degree of confidence 0.5; and so on. Kelly argues that certain conciliatory views should adopt a fine-grained analysis of the possible doxastic attitudes.[8] Nothing I say below depends upon whether there are exactly three possible doxastic attitudes, as a course-grained approach suggests, or more than three, as a fine-grained approach suggests. My argument requires only that there are at least three such attitudes, whatever they happen to be.

The literature on the epistemology of disagreement considers situations in which epistemic peers adopt differing doxastic attitudes toward a proposition after fully disclosing to each other the reasons for their attitude toward that proposition. Christensen defines two people as *epistemic peers* regarding a proposition just in case they have considered roughly the same evidence with respect to whether that proposition is true and they are roughly equally good at responding to that kind of evidence.[9] While there are other definitions available in the literature, this suffices as a working definition. Nothing in my argument hinges upon its correctness. I require only that there are at least two epistemic peers. When there are not at least two peers, my argument does not hold. But since the situations of interest to epistemologists are those in which there is disagreement, and since every disagreement involves at least two peers, this limitation is not significant.

Moreover, all of the situations of interest to epistemologists are ones in which epistemic peers adopt different doxastic attitudes toward the same proposition. I shall say that an attitude a person adopts toward a proposition is ON for that person with respect to that proposition, and that an attitude a person does not adopt toward a proposition is OFF. For example, if the possible doxastic attitudes are coarse-grained and if, regarding the proposition that God exists, believing is the only attitude Aquinas has toward it, then *believing that God exists* is ON for Aquinas while both *disbelieving that God exists* and *withholding judgment about God's existence* are OFF. There is peer disagreement regarding a proposition when the peers have different doxastic attitudes ON toward that proposition.

Regardless of what the possible doxastic attitudes are, each of a person's possible doxastic attitudes toward a proposition is either ON or OFF for that person toward that proposition. But it seems that there are situations in which one attitude can be *less* OFF (or more ON) for a person toward a proposition than another. For example, imagine a theist and atheist discussing whether God exists. Suppose the conversation turns to agnosticism, the view that our available evidence does not warrant either believing or disbelieving that God exists. Further suppose that they

[8] Kelly, "Peer Disagreement and Higher-Order Evidence," 117-118.

[9] Christensen, "Epistemology of Disagreement," 211.

both admit that agnosticism is more plausible than their opponent's view, even though each retains her belief. Then *withholding judgment about God's existence* is less OFF for the atheist than is *believing that God exists*, and it is less OFF for the theist than is *disbelieving that God exists*. Surely this kind of situation is common; but it is a situation in which people's rankings of possible doxastic attitudes have more than two levels. Also consider

> **Ranking.** Lucy is on *Let's Make a Deal*. Lucy will only choose a door when she believes the prize is behind it; otherwise, she will walk away from the game rather than make a choice. Lucy initially believes that the prize is behind the leftmost of three doors, and so she chooses that door. Regardless of which door Monty Hall reveals to contain a goat, Lucy will continue to believe that there is a prize behind one of the two unopened doors, and in fact she will come to believe that the prize is behind the unchosen and unopened door. She will not walk away from the game.

Let the proposition R be: *The prize is behind the rightmost of the three doors*. When Lucy initially chooses the leftmost door, the attitude *disbelieving R* is ON for her, while the attitudes *believing R* and *withholding judgment about R* are OFF. But it seems that, prior to Monty Hall opening one of the two unchosen doors, *believing R* is *less* OFF for Lucy than *withholding judgment about R*. For, at that time, she is more disposed to change from disbelieving R to believing R than she is to change from disbelieving R to withholding judgment about R. When Lucy is disbelieving R, the (non-actual) possible world in which she believes R is closer than the world in which she withholds assent about R. Given this, it seems that when *disbelieving R* is ON for Lucy, *believing R* is less OFF for her than is *withholding judgment about R*. (For similar reasons, it seems that, when Lucy chooses the leftmost door, *believing that the prize is behind the center door* is less OFF for her than is *withholding judgment about whether the prize is behind the center door*.)

The *is less OFF than* relation is obviously *transitive*: for any person S, proposition P, and distinct doxastic attitudes X,Y,Z toward P, whenever X is less OFF for S than is Y and Y is less OFF for S than is Z, X is less OFF for S than is Z. Regarding the Ranking case, transitivity entails that, when Lucy initially chooses the leftmost door, *disbelieving R* is less OFF for her than is *withholding judgment about R*, because any ON attitude is less OFF than any OFF attitude.

Transitivity is an essential presupposition for the Arrovian-style impossibility theorem for conciliatory pure HOE views of peer disagreement. Also essential is a modal claim about rankings of doxastic attitudes toward propositions.

> *Depth*: It is possible that there exists a person S, proposition P, and distinct doxastic attitudes X,Y,Z toward P such that X is less OFF for S than is Y and Y is less OFF for S than is Z.

The Ranking case supports *Depth*. When Lucy initially chooses the leftmost door, *disbelieving R* is less OFF for her than is *believing R* (by virtue of *disbelieving R* being ON) and *believing R* is less OFF for her than is *withholding judgment about R*. *Depth* entails that the *is less OFF than* relation orders people's doxastic attitudes toward propositions in a way that does not necessarily have only two ranking levels.

3. Response Functions and Doxastic Attitude Rankings

Pure HOE views about peer disagreement may be understood as maintaining that there is a function that takes as input information about higher-order evidence about disagreeing peers' doxastic attitudes toward a disputed proposition and yields as output a verdict about the rational response to that disagreement after the peers disclose to each other the (first-order) reasons for their conflicting attitudes. For example, the equal weight view may be understood as maintaining that the following function is correct for the case in which two epistemic peers disagree about some proposition P

$$\text{(EWV):}\quad (C_1 + C_2)/2 = C_R,$$

where C_1 is the credence peer 1 gives to P, C_2 is the credence peer 2 gives to P, and C_R is the credence each peer ought to give to P after full disclosure.[10] Similarly, the extra weight view may be understood as proposing as correct the function

$$\text{(XWV):}\quad (C_1+C_2)/2 + x(C_1-C_2)/2 = C_R,$$

where peer 1 is (indexically) the person adjusting her doxastic attitude and x ($0\leq x\leq 1$) is the amount of extra weight that peer gives to her attitude.

Let us call functions like EWV and XWV *response functions* and information about a peer's doxastic attitudes toward a proposition a *doxastic profile* for that peer. Then pure HOE views may be understood as maintaining that the rational response to peer disagreement is determined by a response function that takes as input the doxastic profiles for all disagreeing peers and yields as output a doxastic profile that those peers ought to have after full disclosure. Conciliatory views may

[10] For an objection and alternative to this way of understanding the equal weight view, see Branden Fitelson and David Jehle, "What is the Equal Weight View?" *Episteme* 6 (2009): 280-293.

be understood as adding that the output of this response function should be some kind of compromise among the profiles taken as input.

Response functions need not be mathematical. Consider, for example, Feldman's split the difference view.[11] According to this view, if one peer believes P and another peer disbelieves P, the rational response to this disagreement after these peers disclose their reasons to each other is for each peer to withhold assent about P. This may be represented as a non-mathematical function f_F, where $B_n(P)$ represents that peer n believes that P:

(SDV): $f_F(B_1(P), B_2(\neg P)) = \neg B(P)\&\neg B(\neg P)$.

While SDV itself has the appearance of a mathematical equation, the function f_F is not mathematical, in the same way that the function $f_\&(P,Q)$ for conjunction-introduction is not mathematical.

The output to a response function need not be a doxastic profile in which there is a *unique* doxastic attitude that disagreeing peers ought to have after full disclosure. Some pure HOE views, like the equal weight view, maintain that there is exactly one doxastic profile all peers ought to have after full disclosure; others, like the extra weight view, allow peers to have different profiles after full disclosure by virtue of advocating *indexical* response functions. There even could be non-indexical response functions that allow more than one doxastic attitude as the rational response to peer disagreement after full disclosure.[12] Accordingly, understanding pure HOE views in terms of response functions is neutral regarding whether, for any given evidential situation, there is only one rational response to peer disagreement after full disclosure in that situation.[13]

Information about the doxastic profiles taken as input for conciliatory pure HOE response functions cannot be *merely* information about which doxastic attitudes happen to be ON for the peers, even though typical presentations of such views give this impression. For there is some reason to think that, if the input were restricted in this way, conciliatory pure HOE views would face insuperable difficulties.

Consider a situation in which two epistemic peers, an atheist and an agnostic, are discussing whether God exists. Suppose that there are three possible doxastic attitudes: believing; disbelieving; withholding assent. Conciliatory views about disagreement entail that, after full disclosure, each peer should change his doxastic

[11] Feldman, "Epistemological Puzzles about Disagreement."

[12] For some suggestions, see Kelly, "Peer Disagreement and Higher-Order Evidence," 120-121.

[13] For further discussion of uniqueness, see Roger White, "Epistemic Permissiveness," *Philosophical Perspectives* 19 (2005): 445-459.

attitude in the direction of the other. But, as Kelly notes, there is no suitable way to do so.[14] Kelly takes this to entail that conciliatory views should adopt a more fine-grained approach to possible doxastic attitudes. But this precludes the problem only if those attitudes are dense, so that there is always another attitude between any distinct doxastic attitudes. For if the attitudes are discrete, conciliatory views flounder in situations where disagreeing peers adopt conflicting attitudes toward a proposition and there is no "middle-ground" attitude available. However, it is extremely unlikely that the doxastic attitudes had by actual people are just as fine-grained as, say, the real numbers. So Kelly's proposal on behalf of conciliatory views preserves their truth at the cost of rendering them inapplicable to the actual world.

Conciliatory views about disagreement can avoid the preceding difficulty without endorsing an unrealistic view about possible doxastic attitudes, by allowing input to response functions to include more than information about which attitudes happen to be ON for the peers after full disclosure. For conciliatory views that are also pure HOE views, this further information must be information about higher-order evidence. The only such information is information about how peers rank possible doxastic attitudes in terms of the *is less OFF than* relation. Fortunately, this solves the problem without the costs of Kelly's proposal.

Consider again the disagreeing atheist and agnostic. The atheist has *disbelieving that God exists* ON, while the agnostic has *withholding judgment about whether God exists* ON. Since their doxastic attitudes differ, the rest of their doxastic profiles must differ as well. For example, perhaps the atheist's profile is such that: *disbelieving that God exists* is less OFF than both *withholding assent that God exists* and *believing that God exists*, while neither of these latter two attitudes is less OFF than the other; and perhaps the agnostic's profile is such that *withholding assent that God exists* is less OFF than both *disbelieving that God exists* and *believing that God exists*, while neither of these latter two attitudes is less OFF than the other. If conciliatory views require only that two disagreeing peers change their doxastic *profiles* toward each other (rather than change the *attitudes* that they happen to have ON) after full disclosure, such views can maintain that the disagreeing peers should change their rankings of attitudes that are OFF. So, for example, in the case of the atheist and agnostic, such a view might maintain that the atheist should adopt a profile in which *disbelieving that God exists* is less OFF than *withholding assent that God exists*, which in turn is less OFF than *believing that God exists*, and that the agnostic should adopt one in which *withholding assent that God exists* is less OFF than *disbelieving that God exists*, which in turn is less OFF than *believing that God exists*. This kind of response to peer disagreement does

[14] Kelly, "Peer Disagreement and Higher-Order Evidence," 117.

not *remove* the disagreement between the atheist and the agnostic; but then, other conciliatory pure HOE views, such as the extra weight view, also allow the disagreement to persist. Since disagreeing peers are guaranteed to have differing doxastic profiles, *some* kind of change among the OFF attitudes for each peer is always possible. Accordingly, conciliatory pure HOE views can avoid the problem Kelly raises without making themselves inapplicable to the actual world, provided that they propose response functions that take as input information about the rankings in peers' doxastic profiles.

Extant conciliatory and pure HOE views of peer disagreement do not consider response functions that take this kind of information as input. Nor, for that matter, do steadfast or mixed evidence views. For this reason, the peer disagreement literature has yet to consider adequacy conditions for such response functions. One *prima-facie* plausible condition is that, for any pair of distinct doxastic attitudes, such functions should yield as output a relative ranking of those attitudes that is independent of changes in peers' doxastic profile rankings for other pairs of attitudes after the peers fully disclose to each other the reasons for their attitudes.

> *IIA*: For any proposition P and any distinct doxastic attitudes X,Y toward P, if some or all peers change their doxastic profiles toward P after full disclosure without changing the relative ranking of X and Y within those profiles, the output of the response function does not change the relative ranking of X and Y.

(*IIA* abbreviates *I*ndependence of *I*rrelevant *A*lternatives.) Consider an abstract situation in which, for some proposition P and doxastic attitudes X,Y, and Z toward P, the output of the response function yields that X should be less OFF than both Y and Z. This output is based upon full disclosure of all evidence among epistemic peers and, perhaps, the doxastic profiles of the peers after this disclosure. The output is either eternally correct for the peers' evidential situation or not. If it is eternally correct, then if some of the peers change their doxastic profiles without acquiring new evidence (or losing available evidence), the output should remain as it was initially, because the peers' evidential situation remains the same. This accords with *IIA*. If the initial output is not eternally correct, the updated output of the response function depends, at least in part, upon the changed doxastic profiles of the peers. The intuition driving *IIA* in this condition is that updates to response function output should be proportionate to changes in peers' doxastic profiles. (If a peer changes the relative ranking of attitudes X and Y but not the relative ranking of X and Z, then if the response function output requires updating, I say that the updating is *proportionate* just if the function's output changes the relative ranking of attitudes X and Y but not the relative ranking of X and Z.) The motivation for this intuition is that, when a peer changes one pairwise ranking of doxastic attitudes but not other pairwise rankings despite acquiring no new evidence (and losing no available evidence), there is no reason that warrants changing any of the other pairwise rankings, because all the initial evidence is the same; and when nothing warrants a change in pairwise rankings that are rational, changing those rankings would be irrational. If, say, there is no reason that warrants changing the rational relative ranking of attitudes X and Z, changing this relative ranking would be irrational, and so the response function's updated output regarding the relative ranking of X and Z should remain unchanged.

4. Constraints on Conciliatory Pure HOE Response Functions

Conciliatory pure HOE views impose two conditions on response functions that make them incompatible with *IIA*. The first is that there is no peer such that that peer's ranking one doxastic attitude as less OFF toward a proposition than another after full disclosure strictly implies that output of the response function ranks the former attitude as less OFF toward that proposition than the latter attitude.

> *Fallibility*: It is not the case that there exists a peer such that, for any proposition P and any distinct doxastic attitudes X,Y toward P, necessarily, whenever that peer ranks X as less OFF than Y after full disclosure, the response function yields as output a ranking in which X is less OFF than Y.

All conciliatory views about disagreement endorse *Fallibility*. If *Fallibility* were false, then there could be a peer disagreement in which at least one party to the dispute is not required, after full disclosure, to change his attitude toward the disputed proposition to one that is closer to the attitudes of those peers with which he disagrees. But, according to conciliatory views, such change is always required of all peers.

The second condition on response functions for conciliatory pure HOE views concerns situations in which all peers have the same pairwise ranking of distinct possible doxastic attitudes toward a proposition after full disclosure.

> *Unanimity*:For any proposition P and any distinct doxastic attitudes X,Y toward P, if all peers rank X as less OFF toward P than Y after full disclosure, the response function yields as output a ranking in which X is less OFF than Y.

For example, according to *Unanimity*, if everyone flat-out believes that the continuum hypothesis is true after fully disclosing to each other the reasons for their belief, the rational response to this situation is to rank *believing the continuum hypothesis* as less OFF than both *disbelieving the continuum hypothesis* and *withholding judgment about the continuum hypothesis*. If *Unanimity* is false, then there is some proposition P and distinct attitudes X,Y such that, although all peers rank X as less OFF toward P than Y after full disclosure, those peers ought to change their doxastic profiles so as to not rank X as less OFF toward P than Y. However, according to pure HOE views, no peer in such a situation has any evidence to support changing her original assessment of the evidence for P, and so no peer ought to change her original doxastic profile after full disclosure.

All pure HOE views endorse *Unanimity*. For example, according to both the equal weight view and the extra weight view, if everyone has a credence of 0.9 toward P after full disclosure, having a credence of 0.9 toward P is the rational attitude to have. (Strictly speaking, pure HOE views do not apply to cases of unanimous peer *agreement*; but they should extend naturally to such cases in a way that validates *Unanimity*.) The falsity of *Unanimity* opens the possibility that, even if everyone has a credence of 0.9 toward P after full disclosure, that is not the rational credence to have, because some other credence should be less OFF toward P. But if everyone's evidence leads them to have a credence of 0.9 toward P after full disclosure, no one has reason to revise their credence. Also, consider

> **Ranking 2.** Before Monty Hall opens the center door for Lucy, Lucy consults Marilyn, her off-stage friend. Lucy discovers that Marilyn also believes that the prize is behind the leftmost door, that Marilyn will continues to believe that the prize is behind *some* door no matter which one Monty opens, and that Marilyn will come to believe that the prize is behind the unchosen and unopened door after Monty opens a door.

Before Lucy consults with Marilyn to discuss each other's reasoning, both women rank *believing R* as less OFF than *withholding judgment about R*. (R, recall, is the proposition that *the prize is behind the rightmost door*.) After consulting with each other, neither acquires any higher-order evidence to support revising this ranking. In accordance with *Unanimity*, any pure HOE view must thereby entail that the rational response to the women sharing their reasoning with each other is for both women to retain their original ranking of *believing R* as less OFF than *withholding judgment about R*.

5. An Arrovian-Style Impossibility Theorem

Unanimity and *IIA* jointly entail that *Fallibility* is false. I shall call this result *Arrow's Epistemological Theorem*. Since conciliatory pure HOE views entail both *Unanimity* and *Fallibility*, and since the motivation for *IIA* is that updates to response function outputs after full disclosure should be proportionate to changes epistemic peers make to their doxastic profiles after full disclosure (if, indeed, such outputs should be updated at all), this theorem amounts to the claim that conciliatory pure HOE views demand disproportionate updates of response function output when peers change their doxastic profiles after full disclosure despite acquiring no new evidence (and losing no available evidence).

The proof of Arrow's Epistemological Theorem, following Geanakoplos, involves three steps.[15] The first shows that, for *any* doxastic attitude Y, if, after full disclosure, everyone in a peer group ranks Y as either not less OFF than anything else or less OFF than everything else, then the response function must rank Y as either not less OFF as anything else or less OFF than everything else. The second shows that, for a *particular* doxastic attitude Y, there is someone in the peer group who is *infallible* with respect to all pairwise rankings not involving Y. The third step shows that this same person must be infallible with respect to all pairwise rankings, regardless of whether they involve Y.

Step 1. Consider a situation in which, after everyone has disclosed to one another the reasons for their attitudes toward some arbitrary proposition, all epistemic peers have doxastic profiles that rank some arbitrary doxastic attitude Y toward that proposition as *either* not less OFF than any other attitude or less OFF than all other attitudes: after full disclosure, everyone's profile has either Y ON and other attitudes OFF, or Y the most OFF of all attitudes. (This situation might be one in which half of the peers rank Y as not less OFF than any other attitude, while the other half rank Y as less OFF than all other attitudes.) *IIA* and *Unanimity* entail

> *Extremal Lemma*: For any doxastic attitude Y toward a proposition and any peer set of doxastic profiles for that proposition, whenever every peer ranks Y as either not less OFF than any other attitudes or less OFF than all other attitudes after full disclosure, the output of any response function must either rank Y as not less OFF than any other attitude or else rank Y as less OFF than all other attitudes.

Suppose, for *reductio*, that the response function does not rank Y in either of these ways. Then there are attitudes X,Z such that the response function yields, as output, that X should be less OFF than Y and Y should be less OFF than Z. Now suppose that, for whatever arbitrary reason and despite no change in available evidence, every peer's doxastic profile changes so that each person ranks Z as less OFF than X while not changing their pairwise rankings involving Y. Then *IIA* entails that the response function continues to yield, as output, that X should be less OFF than Y and Y should be less OFF than Z. Transitivity of the *is less OFF than* relation entails that this function yields that X should be less OFF than Z. However, *Unanimity* entails that the function yields that Z should be less OFF than X. Discharging the contradiction and completing the *reductio* establishes the lemma.

Step 2. Next, consider a particular doxastic attitude Y toward a proposition and a situation in which all peers have doxastic profiles that rank Y as more OFF

[15] Geanakoplos, "Three Brief Proofs," 212-213.

than all other attitudes after full disclosure (otherwise the rankings in the peer profiles are arbitrary). Call this Situation 1. Imagine that, for whatever arbitrary reason and despite no change in available evidence, each of N peers successively changes her profile so that Y goes from being ranked as *more* OFF than all other attitudes to being ranked as *less* OFF than all other attitudes. Let Situation N be the situation, after full disclosure, in which all peers have doxastic profiles that rank Y as less OFF than all other attitudes. In Situation 1, *Unanimity* entails that the output of the response function should rank Y as more OFF than all other attitudes. The Extremal Lemma entails that, for every situation between Situation 1 and Situation N, the response function should either rank Y as more OFF than all other attitudes or else rank Y as less OFF than all other attitudes. In Situation N, *Unanimity* entails that the output of the response function should rank Y as less OFF than all other attitudes. Clearly, there must exist a peer, n*, whose profile change causes a change in the output of the response function.

Let Situation A be one in which this n* has a doxastic profile that ranks Y as more OFF than all other attitudes, and let Situation B be like Situation A except that n* has changed to have a profile that ranks Y as less OFF than all other attitudes. Then the output of the response function in Situation A should rank Y as more OFF than all other attitudes; and in Situation B, it should rank Y as less OFF than all other attitudes. Consider two arbitrary doxastic attitudes X,Z, each distinct from Y, and construct an arbitrary Situation C from Situation B that satisfies the following conditions:

- the profiles for peers 1 through n*-1 rank Y as less OFF than any other attitude,
- the profiles for peers n*+1 through N rank Y as more OFF than any other attitude, and
- the profile for n* ranks X as less OFF than Y and Y as less OFF than Z.

IIA entails that output of the response function regarding the relative ranking of X and Y for Situation C should be the same as it is for Situation A. Given the relation between Situation A and Situation B, Situation A and Situation C have the same pairwise rankings of X and Y for all peer profiles. Since, in Situation A, the output of the response function is that X should be less OFF than Y, *IIA* entails that this is the output of the response function in Situation C as well. Similarly, given the relation between Situation B and Situation C, those situations have same pairwise rankings of Y and Z for all peer profiles. Since, in Situation B, the output of the response function is that Y should be less OFF than Z, *IIA* entails that this is the output of the response function in Situation C as well. Transitivity of the *is less OFF* *than* relation thereby entails that, in Situation C, the output of the response function should rank X as less OFF than Z.

A similar argument shows that if, in Situation C, the doxastic profile for n* were to rank Z as less OFF than Y and Y as less OFF than X, the output of the response function in Situation C would rank Z as less OFF than X. Hence, for a *particular* doxastic attitude Y, there is an n* in the peer group who is infallible with respect to all pairwise rankings not involving Y, in the sense that this person determines the response function's output for how those alternatives should be ranked. A similar argument, considering a *different* particular doxastic attitude Z, shows that there is also a person, n**, in the peer group who is infallible with respect to all pairwise rankings not involving Z.

Step 3. Suppose, for *reductio*, that n* is not the same person as n**. Then n* cannot affect the response function's output regarding the relative ranking of alternatives X and Y, because n** determines that output. Yet clearly sometimes n* does affect this output, as with Situations A and B. Hence, n*=n**. Similar arguments show that n* determines the response function's output for *all* rankings, and this amounts to *Fallibility* being false.

Therefore, if *Unanimity* and *Fallibility* are true, *IIA* is false. This is Arrow's Epistemological Theorem, and it places a burden on advocates of conciliatory pure HOE views.

If they accept the theorem, their burden is to show that, when updates to response function outputs after full disclosure are *not* proportionate to changes epistemic peers make to their doxastic profiles after full disclosure, the updated outputs continue to capture rational responses to evidential situations among epistemic peers. If they reject the theorem, their burden is to show that some background presupposition for the theorem fails.

I maintain that updates to response function outputs are rational only if they are proportionate, so that any view that denies *IIA* is false. So far as I know, the extant literature on peer disagreement does not provide an argument to the contrary. Accordingly, if Arrow's Epistemological Theorem is sound, it shows that no pure HOE view can be conciliatory. For pure HOE views endorse both *Unanimity*, conciliatory views endorse *Fallibility*, and the theorem shows that *Unanimity* and *Fallibility* jointly entail that *IIA* is false.

6. Prospects for Avoiding Arrow's Epistemological Theorem

If no pure HOE view can be conciliatory, one of the most popular views about peer disagreement, the equal weight view, must be mistaken. Since many epistemologists have strong intuitions that something like the equal weight view must be true, it is worth considering some options for rejecting Arrow's Epistemological Theorem. The theorem, after all, requires several background presuppositions, and if one of these were to be false, the theorem would not be sound.[16] I shall consider the prospects for denying four such presuppositions, arguing that each prospect is unpalatable for those who accept views about peer disagreement that are both conciliatory and pure HOE.

An advocate of a conciliatory pure HOE view might object that, even if such a view may be understood as maintaining that there is a function that takes as input information about higher-order evidence about disagreeing peers' doxastic attitudes toward a disputed proposition after the peers disclose to each other the reasons for their conflicting attitudes and yields as output a verdict about the rational response to that disagreement, the output of this function is not a ranking of doxastic attitudes in terms of the *is less OFF than* relation. Instead, the objection might go, the output of a response function is merely information about which particular attitude(s) peers ought to adopt toward a proposition after full disclosure. This is output about which attitude(s) should be the least OFF one(s). However, even if this is correct, analogues of Arrow's Epistemological Theorem hold under reasonable conditions.[17] So this option does not seem promising.

Rather than focusing on outputs of response functions, an advocate might focus on inputs, objecting that response functions need take as input *only* information about which peer doxastic attitudes happen to be ON after the peers disclose to each other the reasons for their conflicting attitudes. After all, the argument against this understanding of response function input relies upon a special kind of case, namely, one in which disagreeing peers adopt conflicting doxastic attitudes toward a proposition after full disclosure and there is no "middle-ground" doxastic attitude for them toward which they can move. That conciliatory pure HOE views fail to handle this kind of case does not show that they do not handle *any* kind of peer disagreement. Hence, this objection goes, even if Arrow's Epistemological Theorem shows that conciliatory pure HOE views are false when applied to a special kind of case, it does not show that such views are false more generally.

While this objection is cogent, it rescues conciliatory pure HOE views about disagreement from refutation at the cost of making them unattractively ad hoc. If advocates of conciliatory pure HOE views opt to restrict the range of cases to which such views apply, then, in the special cases, either some peer need not change her

[16] I am indebted to the discussion of some of these presuppositions in Samir Okasha, "Theory Choice and Social Choice: Kuhn versus Arrow," *Mind* 120 (2011): 83-115.

[17] See Austen-Smith and Banks, *Positive Political Theory I*, 49-52.

doxastic attitude toward the others or else first-order evidence helps to determine how the peers should change their attitudes. But there does not seem to be a principled reason for allowing that a steadfast response is rational when there is no "middle-ground" attitude and yet denying that a steadfast response is rational when there is, because facts about how many possible doxastic attitudes happen to be available between two peers' conflicting attitudes are not facts about higher-order evidence (thereby violating the spirit, if not the letter, of pure HOE views), and because such facts do not seem to be relevant to the rationality of a response to peer disagreement. Moreover, maintaining that there is a default doxastic attitude, such as withholding assent, removes the appearance of adhockery by virtue of not being a conciliatory view. For if, say, the proper rational response to disagreement after full disclosure between a theist and an agnostic is for both to withhold assent about whether God exists, the agnostic's doxastic attitude remains unchanged.

Perhaps, however, advocates of conciliatory pure HOE views can avoid the charge of adhockery by denying that the special cases pose any problem at all. The argument that they do depends upon the claim that possible doxastic attitudes for actual people are not dense. But, one might object, an advocate of a conciliatory pure HOE view need not be moved by this contingent fact, because the claim to the contrary may be understood as an idealization, and idealized theories do not merit any special concern. For example, even though the equation of motion for the simple pendulum is idealized by virtue of treating the pendulum bob as a point-mass particle and the pendulum string as perfectly rigid (among other things), the equation remains useful and legitimate to use for certain situations in which these idealizing conditions do not obtain.

There is something correct about this objection. Idealized theories often are not particularly worrisome. Nonetheless, the objection is flawed. The idealizations that do not cause concern are *controllable*: there is some way to take into account the distorting effects of the idealization.[18] This accounting might involve removing the idealization, showing that its effect on the theory is negligible, and so on.[19] However, the density idealization is not controllable, because response functions for density-idealized conciliatory pure HOE views produce outputs that their counterpart non-density-idealized response functions deem to be impossible. There is no way to remove the density idealization, or to estimate the idealization's effect

[18] See Lawrence Sklar, *Theory and Truth: Philosophical Critique within Foundational Science* (New York: Oxford University Press, 2000), 44, 61-70.

[19] For example, see Ronald Laymon, "Idealization, Explanation, and Confirmation," *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Volume One (1980): 336-350.

because, without the idealization, response function outputs for conciliatory pure HOE views are either incorrect or inapplicable to the actual world. In this respect, the density-idealization is akin to the idealization of systems as having infinitely many particles in statistical mechanical accounts of phase transitions.[20] When idealizations are uncontrollable, it is not clear that theories which rely upon them have any applicability to the real world. If they do not, such theories might be true of some idealized situations, but they are false of real ones.

A fourth way for an advocate of a conciliatory pure HOE view to avoid Arrow's Epistemological Theorem is to maintain that *Depth* is false. If one doxastic attitude can be less OFF than another only when the former is ON and the latter is OFF, the proof of Arrow's Epistemological Theorem fails. However, *Depth* is an extremely weak claim. Its truth is compatible with all actual people's rankings of doxastic attitudes being such that one attitude is less OFF than another only when the former is ON and the latter is OFF. Even if thinking of doxastic attitudes as being ON or OFF and ranking doxastic attitudes in terms of the *is less OFF than relation* is new, this novelty alone does not support the strong modal claim that *Depth* is false, especially when the Ranking case provides at least some evidence to the contrary.

The responses to the preceding objections suggest that conciliatory pure HOE views about peer disagreement are false if they cannot avoid Arrow's Epistemological Theorem (by virtue of violating *IIA*) and that they can avoid Arrow's Epistemological Theorem only by virtue of being unacceptably ad hoc. There are other views about peer disagreement that can accept the theorem without being ad hoc and without violating *IIA*. But these are unpalatable for views that are both conciliatory and pure HOE, because they involve adopting views that are either steadfast or mixed. For example, consider

> **Extreme.** Two rationally competent peers mistake the import of a shared body of evidence regarding hypothesis H. In response to the evidence, one peer gives credence 0.7 to H and the other gives it 0.9. However, the evidence *in fact* supports only the credence 0.3 for H.[21]

Kelly takes this kind of case, in which disagreeing peers radically misevaluate the import of their evidence, to show that pure HOE views are incorrect. Suppose he is right. But suppose that these kinds of cases support a view according to which,

[20] See Craig Callender, "Taking Thermodynamics Too Seriously," *Studies in History and Philosophy of Science Part B* 32 (2001): 539-553; Chuang Liu, "Infinite Systems in SM Explanations: Thermodynamics Limit, Renormalization (Semi-) Groups, and Irreversibility," *Philosophy of Science* 68 (2001): S325-S344.

[21] This adopts Case 5 in Kelly, "Peer Disagreement and Higher-Order Evidence," 125-126.

when all peers have the same attitude toward a proposition, the rational response to the evidence for that proposition is a function of first-order evidence *only*. This is a mixed view, and it entails that *Unanimity* is false. For even if everyone were to mistake the import of the evidence for a hypothesis and adopt the same incorrect credence toward that hypothesis after full disclosure, the rational response to the evidence would not be to adopt that particular mistaken credence.

7. Concluding Remarks

Whether a person has a particular doxastic attitude toward a proposition is not an all-or-nothing affair. For there are situations in which one doxastic attitude for a person toward a proposition can be less OFF than another attitude of that person toward the same proposition (see Section 2). The extant literature on the epistemology of peer disagreement overlooks this kind of depth in people's doxastic attitudes. But acknowledging this depth allows conciliatory pure HOE views of peer disagreement to avoid certain difficulties, by virtue of denying that the information about doxastic profiles of epistemic peers taken as input by response functions for such views is merely information about which doxastic attitudes happen to be ON for those peers (see Section 3).

An adequacy condition for response functions that take as input more information than information concerning which doxastic attitudes happen to be ON for epistemic peers is *IIA* (Independence of Irrelevant Alternatives): for any pair of distinct doxastic attitudes toward a proposition, if some epistemic peers change their doxastic profiles toward that proposition after full disclosure, without changing the relative ranking of those doxastic attitudes, the output of the response function does not change the relative ranking of those attitudes either. This condition ensures that updates to response function outputs do not change relative rankings of doxastic attitudes without reason (see Section 3). Conciliatory pure HOE views impose additional constraints on response functions (see Section 4). Yet, according to Arrow's Epistemological Theorem, these constraints are jointly incompatible with *IIA* (see Section 5). Accordingly, given *IIA*, if Arrow's Epistemological Theorem is sound, no pure HOE view of peer disagreement can be conciliatory and, in particular, the popular equal weight view is mistaken.

While there are ways to avoid Arrow's Epistemological Theorem, none of them should be appealing to advocates of conciliatory pure HOE views (see Section 6). Restrictions on the output of response functions succumb to analogues of the theorem. Restricting the inputs of response functions makes conciliatory pure HOE views either ad hoc or inapplicable to real cases, thereby preventing them from providing a general answer to the question of what disagreeing peers ought to do.

Finally, rejecting certain constraints on response functions themselves involves adopting views about peer disagreement that are either steadfast or mixed.[22]

[22] I thank George Schumm and participants at the 2011 Omaha Epistemology Workshop for helpful comments on earlier versions of this paper.

THE TEMPORAL GENERALITY PROBLEM

Brian WEATHERSON

ABSTRACT: The traditional generality problem for process reliabilism concerns the difficulty in identifying each belief forming process with a particular kind of process. That identification is necessary since individual belief forming processes are typically of many kinds, and those kinds may vary in reliability. I raise a new kind of generality problem, one which turns on the difficulty of identifying beliefs with processes by which they were formed. This problem arises because individual beliefs may be the culmination of overlapping processes of distinct lengths, and these processes may differ in reliability. I illustrate the force of this problem with a discussion of recent work on the bootstrapping problem.

KEYWORDS: reliabilism, generality, bootstrapping

1. Two Kinds of Generality Problem

The generality problem is a well-known problem for process reliabilist theories of justification.[1] Here's how the problem usually gets started. In the first instance, token processes of belief formation are not themselves reliable or unreliable. Rather, it is *types* of processes of belief formation that are reliable or unreliable. But any token process is an instance of many different types. And these types may differ in reliability.

For instance, imagine I read in the satirical newspaper *The Onion* that Barack Obama is the president. On this basis, I come to believe that Barack Obama is the president. The process I have used to form this belief is an instance of each of these types.

1. Coming to believe that Barack Obama is the president;
2. Believing something because it was written in *The Onion*; and
3. Believing something because it was written in a newspaper.

The first type of process is very reliable, at least in 2012. The second is highly unreliable, and the third is very reliable. So should we say that the token process I

[1] On process reliabilism, see Alvin Goldman, "What is Justified Belief," in *Justification and Knowledge*, ed. George Pappas (Dordrecht: Reidel, 1979), 1-23. On the generality problem, see Richard Feldman, "Reliability and Justification," *Monist* 68 (1985): 159-174; Earl Conee and Richard Feldman, "The Generality Problem for Reliabilism," *Philosophical Studies* 89 (1998): 1-29.

used was reliable or unreliable? More generally, is there a principled way to map token processes to types of process in a way that lets us systematically say whether a particular process is reliable or not? Critics of reliabilism argue that there is not.

As I said, this problem has been around for quite a while, but I don't think the full force of the problem has been appreciated. Reliabilism is a theory about whether a belief is justified or unjustified. But to determine whether the belief is justified, we step back from the belief itself in two respects. First, we look not to the belief, but to the token process of belief formation from which it results. Second, we look not just to that process, but to kinds of processes of which it is an instant. When carrying this out, we need to make the following two mappings.

- Belief → Token process of belief formation;
- Token process of belief formation → Type of process of belief formation

The traditional point of the generality problem is that the second of these mappings is one-many, not one-one. Each token process is associated with many, many types of processes. But what hasn't been sufficiently appreciated is that the first mapping is one-many as well. And this generates a new, and potentially harder, form of the generality problem.

That the first mapping is one-many isn't because of any special properties of beliefs. Typically, an event is the conclusion of more than one process. Imagine that I travel from Michigan to New York to see a friend. I conclude this journey by walking to the friend's apartment. With the last step I take, I conclude several processes. These include:

1. Walking from the subway station to the apartment;
2. Travelling by public transit from the airport to the apartment; and
3. Travelling from Michigan to my friend's apartment.

It is possible that one of these is a quite reliable process, while the others are not. If I am good at navigating the Manhattan street grid by foot, but poor at making it to the airport on time, then process one will be a highly reliable process, while process three will not. So should we say that my arrival at my friend's apartment was the result of a reliable process or not? The best reply to that question is to point out that it is ill formed. Given that I made it to the nearest subway station, I used a reliable process to traverse the last few blocks. But the longer process I used was not as reliable.

This raises a conceptual worry for process reliabilist theories. If there is no such thing as the reliability of a conclusion, but only the reliability of a process of getting from one or other starting point to that conclusion, then it seems that in identifying the justifiedness of a belief with *the* reliability of the process used to generate it, we commit a kind of category mistake. Note that this problem would persist even if we had a one-one mapping from token processes to epistemologically relevant types of processes that would let us solve the traditional form of the generality problem. We would still need a way of saying which of the many processes which terminate in a belief is the epistemologically relevant one. I don't think there's any reason to think there is a good answer to this question. I call this the *Temporal* Generality Problem, because the different processes that culminate in a belief are typically of different durations.

2. Can the Problems be Solved Simultaneously?

I've argued in the previous section that in theory the Temporal Generality Problem is distinct from the traditional version of the generality problem. But one might think that in practice a solution to the latter will solve problems to do with the former. Consider the following three step process.

1. I hear an astrologer say that Napoleon Bonaparte will win the 2013 US Presidential election.
2. I form the belief that Napoleon Bonaparte will win the 2013 US Presidential election.
3. I deduce that there will be a US Presidential election in 2013.

The process by which I got from 2 to 3 is, on the face of it, highly reliable. Assuming that I'm a mostly sensible person, coming to believe obvious logical consequences of my prior beliefs is a highly reliable process. Yet clearly the process that runs from 1 to 3, i.e., the process of believing obvious logical consequences of the contents of astrological predictions, is not a reliable process. So, one might ask, is the resultant belief justified, because it is formed by the reliable process that runs from 2 to 3, or unjustified, because it is formed by the unreliable process that runs from 1 to 3?

Clearly, this is a false dilemma. The salient kind of process I'm using between 2 and 3 is not *believe obvious logical consequences of a belief*, but *believe obvious logical consequences of a belief* ***formed by an unreliable process***. Once we identify the kind of process used at the last stage correctly, we can see that the unreliability of the whole process causes the process used at the last stage to be unreliable.

We might even get cases that go the other way. There are plenty of occasions in science where scientists use mathematical techniques which cannot be made rigorous, and idealisations that cannot easily be replaced with approximations, or with any other statement known to be true.[2] If we looked at such a step in isolation,

[2] On non-rigorous techniques, see Kevin Davey, "Is Mathematical Rigor Necessary in Physics?" *British Journal for the Philosophy of Science* 54 (2003): 439-463, doi:10.1093/bjps/54.3.4392003;

we would possibly think that it is an unreliable step, even though it is part of a longer, reliable process. But the fact that it is part of a reliable process matters. In particular, it matters to the way we identify the step the scientist is using with a larger kind of inferential processes. That kind won't involve, for instance, all instances of reasoning from false premises, or of reasoning with incoherent mathematical models. Rather, it will just include the kind of reasoning that is licenced by the norms of the science that the scientist is participating in, and that kind might be a very reliable kind of process.

But there is one very special case where I think this kind of solution to the Temporal Generality Problem will not work. It concerns the way in which a reliabilist will try and solve the bootstrapping problem, as developed by Stewart Cohen[3] and Jonathan Vogel.[4] We'll turn next to that problem.

3. Generality and Bootstrapping

Hilary Kornblith[5] has proposed that looking at processes of longer duration generates a reliabilist solution to the bootstrapping problem. I'm going to argue that Kornblith's solution, which I agree is the kind of thing a reliabilist should say, in fact shows that the Temporal Generality Problem is a distinct kind of generality problem, and perhaps a much harder problem than the traditional generality problem.

Let's start with a very abstract version of the bootstrapping problem. Assume device *D* is highly reliable, and *S* trusts device *D* without antecedently knowing that it is reliable. Then the following sequence of events takes place.

- At t_0, *S* sees that device *D* says that *p*.
- At t_1, *S* forms the belief that *D* says at t_0 that *p* on the basis of this perception.[6]
- At t_2, *S* forms the belief that *p*, on the basis that the machine says so.

on idealisations, see Kevin Davey, "Idealizations and Contextualism in Physics," *Philosophy of Science* 78 (2011): 16-38. doi:10.1086/6580932011).

[3] Stewart Cohen, "Basic Knowledge and the Problem of Easy Knowledge," *Philosophy and Phenomenological Research* 65 (2002): 309-329.

[4] Jonathan Vogel, "Reliabilism Leveled," *Journal of Philosophy* 97 (2000): 602-623.

[5] Hilary Kornblith, "A reliabilist solution to the problem of promiscuous bootstrapping," *Analysis* 69 (2009): 263-267. doi:10.1093/analys/anp012.

[6] On some theories of perception, it might be that t_0□=□t_1, since perception involves belief formation. I don't mean to rule those theories out; the notation here is meant to be consistent with the hypothesis that t_0□=□t_1.

- At t_3, *S* forms the belief that the machine is accurate at t_0, on the basis of her last two beliefs.

What should a reliabilist say about all this? Well, the process that runs from t_0 to t_1, the process of believing machine readings are as they appear, looks pretty reliable, so the belief formed at t_1 looks pretty reliable. And the process that runs from t_1 to t_2, i.e., the process of believing that things are as machine *D* says they are, also looks pretty reliable, so that belief looks pretty reliable. And the process that runs from t_2 to t_3, i.e., the process of drawing obvious logical consequences from beliefs formed by reliable processes, also looks pretty reliable. It's true that at t_2, *S* doesn't know she's using a reliable process. And hence at t_3, *S* doesn't know that this is the kind of process that she's using. But none of this should matter to an externalist like the reliabilist, since they think what matters is actual reliability, not known reliability.

But there are two problems lurking in the vicinity. First, many people think that it is very bizarre that *S* can form a justified belief that *D* is accurate at t_0 on the basis of simply looking at *D*. That's the intuition behind the bootstrapping problem. Second, the case looks like an instance of the Temporal Generality Problem. The two problems are related. Kornblith's solution to the bootstrapping problem is to insist that the process used is in fact *unreliable*. What he means to draw our attention to is that the process which runs from t_0 to t_3 is unreliable. And he's right. That looks like a process of determining whether a machine is accurate by simply looking at the machine and trusting it. Of course, there are several other ways we could classify the process used, but Kornblith argues that this is the best classification, and I think he's right. And if he is right, then we have part of a solution to the bootstrapping problem.

But if Kornblith is right, then we pretty clearly also have a nasty instance of the Temporal Generality Problem. Because now it looks like a chain of three reliable processes, those that run from t_0 to t_1, from t_1 to t_2, and from t_2 to t_3, collectively form an unreliable process. The belief that is formed at t_3 is the culmination of two processes; a reliable one that runs from t_2 to t_3, and an unreliable one that runs from t_0 to t_3. If a belief is justified iff it is the outcome of a reliable process, and unjustified iff it is the outcome of an unreliable process, then the belief is both justified and unjustified, which is a contradiction.

How could the reliabilist escape this problem? I can see only two ways out. One is to say that the process that runs from t_0 to t_3 is in fact a reliable process. But that's to fall back into the bootstrapping problem. And in any case, it seems absurd, since that process really does look like a process of determining whether a machine is reliable by simply looking at it. The other is to say that the process that runs from

t_2 to t_3 is unreliable. To do that, we'd need to come up with a natural kind of process which is unreliable, and which this process instantiates. This does not look easy. I'm not going to insist this couldn't be done, but I'll end by noting three challenges that stand in the way of getting it done, and which seem pretty formidible.

First, if we say the process that runs from t_2 to t_3 is unreliable, then we are putting general restrictions on how we can obtain knowledge by deductive inference. As John Hawthorne argues,[7] any such restrictions will be hard to motivate.

Second, the restrictions will have to be fairly sweeping to cover the range of conclusions that, intuitively, cannot be drawn through this kind of reasoning. Imagine a variant on the above example where at t_3, *S* concludes that either *D* is accurate at t_0 or it will snow tomorrow. That's entailed, obviously, by what she knows at t_2. And yet the process of getting from t_0 to that conclusion seems unreliable. So we can't simply say that what's ruled out are cases where the agent draws a conclusion that is simply about *D*.

Third, the classification of the process that runs from t_2 to t_3 must not merely fail to be ad hoc, it must plausibly be the most natural classification available. And yet it seems there is one very natural classification that is not available, namely the classification of the process as an instance of deduction from known premises, or from premises arrived at by highly reliable processes.

So the challenge this problem raises for reliabilism is substantial. I don't mean to say it is a knock-down drawn-out refutation; philosophical arguments rarely are. But it does add a new dimension to the generality problem, and as we've seen in the last few paragraphs, put some new constraints on solutions to the old version of the generality problem.

[7] John Hawthorne, "The Case for Closure," in *Contemporary Debates in Epistemology*, eds. Matthias Steup and Ernest Sosa (Malden: Blackwell, 2005), 26-43.

DEBATE

ON EPISTEMIC ABSTEMIOUSNESS AND DIACHRONIC NORMS: A REPLY TO BUNDY

Scott AIKIN, Michael HARBOUR,
Jonathan NEUFELD, Robert TALISSE

ABSTRACT: In "On Epistemic Abstemiousness," Alex Bundy has advanced his criticism of our view that the Principle of Suspension yields serious diachronic irrationality. Here, we defend the diachronic perspective on epistemic norms and clarify how we think the diachronic consequences follow.

KEYWORDS: disagreement, epistemic abstemiousness, epistemic martyrdom, epistemic conversion

Many thanks to Alex Bundy for his replies[1] to our work[2] and to the editors of *Logos and Episteme* for the opportunity to continue this discussion. In outline, the dialectic stands as follows. We've argued that there are reasons to reject what we've called *the Principle of Suspension* (PS), which runs roughly that *if S is aware that an epistemic peer disagrees with S regarding* p, *S should suspend judgment regarding* p. These reasons arise from our tale of Betty's epistemic journey, wherein she follows PS by first suspending judgment regarding *p* with a disagreeing Alf. Alf's position is improved by this, as he no longer has dissenters. In light of this, Betty now has new evidence for Alf's view, and so must come to agree with him. She may object to Alf's dogmatism on the basis of PS, but if Alf rejects PS, then she is, again, relegated to suspending judgment, not objecting. The trouble, as we saw it, was that PS seems acceptable enough as a *synchronic* epistemic rule, but yields intellectual chaos *diachronically.* Bundy's objections have consistently been (I) that the diachronic social consequences of PS are not relevant considerations for its acceptance, (II) that PS is not the operative principle in yielding Betty's conversion, and (III) that Betty

[1] Alex Bundy, "In Defense of Epistemic Abstemiousness," *Logos & Episteme* II, 2 (2011): 287–92, "On Epistemic Abstemiousness: A Reply to Aikin, Harbour, Neufeld, and Talisse," *Logos and Episteme* II, 4 (2011): 619-624.

[2] Scott Aikin, Michael Harbour, Jonathan Neufeld, and Robert B. Talisse,"Epistemic Abstainers, Epistemic Martyrs, and Epistemic Converts," *Logos & Episteme* I, 2 (2010): 211–9, "On Epistemic Abstemiousness: A Reply to Bundy," *Logos & Episteme* II, 3 (2011): 425–8.

has an out: to hold that Alf is not a peer, because he does not accept or abide by PS. We still think we're on the right track with the argument, but there are some details to clarify.

I

Our main concern is that dialectical-epistemic norms that cannot reasonably be applied iteratedly over time (as a discussion or debate unfolds) have reasons against them. PS is such a norm. It, again, is a norm that expresses a proper concern for what evidence one has, and it takes the attitudes of competent peers as relevant. But when PS is put into motion, we believe it yields social irrationality. Our tale is one of epistemic free-riding and its consequent hazards, as the lesson is that dogmatism pays in contexts of abstemious interlocutors. The consequences that concern us, then, are third-personal and diachronic consequences. For sure, Bundy is right that "the other- and future-regarding notion of rationality … is not the one in question in the debates regarding the appropriate way to respond to disagreement with a peer." And he's right that the norms directed to "now having true beliefs and not having false ones is worthy of study."[3] But what is the fit between these two observations? Bundy's case is that our diachronic and social consequences aren't relevant to the debate, because the debate is about synchronic first personal issues.

Perhaps our reply on this will be too metaphilosophical to cut much ice, but here goes. Compartmentalizing a research program in this fashion is a bad idea, especially when the fact is that those diachronically surveyed *futures* will be *nows* soon enough. Moreover, it seems that if having and understanding the truth is the goal, and if we can show that following a rule like PS impedes that goal over time, that surely is relevant. Our case, we think, is analogous to the person who, in striving to be frugal, buys only the smallest tubes of toothpaste, and thereby spends, in each case of purchasing toothpaste, the least. But over time, this is not so frugal, as toothpaste in those little tubes costs more per ounce, and so it is a better policy in the long run to buy in the big tube. Looking at ourselves only as time-slices is a bad way to knock about in the grocery, and it's bad for epistemology. There, we said it. Now, the fact that most of the folks working on the disagreement problem are exclusively synchronic epistemologists is curious, but of no matter. This is then evidence that we've got a new consideration. Whether Bundy or any of the others see this point as worthy of consideration is on the same level as whether our friend

[3] Bundy, "On Epistemic Abstemiousness," 621.

holds a tiny tube of toothpaste under our noses and insists that all he wants to do is *save money now*. Alright, we concede, be that way.

II

In his first reply, Bundy argued that PS wasn't the principle that yields Alf's improved position or Betty's conversion. We'd interpreted Bundy's counter-argument that Alf's case is not one of applying PS but one of double-counting Alf's view as evidence. We'd argued that such double-counting can work as additional evidence. Our example was that of going back and trying a joke again, but it could also be of, say, counting the pennies in your change basket again, too. But Bundy holds that his case did not depend on challenging the double-counting as vicious. Fair enough, but now the question is *what principle other than PS* is the one that yields the conclusion. Here we're unsure how to respond, because Bundy hasn't proposed an alternative principle.

But there is something to Bundy's challenge, even if he hasn't given us the full-blown version of it. Here, we think, can be the main challenge: PS is an other-regarding epistemic norm that only has the requirement of *suspension* as an output, not endorsement. There needs to be another norm, a cousin to PS, to yield Alf's improved confidence and Betty's conversion, because those two cases are ones of *endorsement*. So something along the lines of the following cluster of conditionals is required:

> (1) If S has a peer that disagrees regarding p, S should suspend judgment; (2) if S has a peer that tends toward agreement that p, S has increased support for p; and (3) if S has no view regarding p but a peer holds that p, S has increased evidence that p.

We'd argued that these come as a family for the following reason: if a peer's beliefs count enough to function as defeaters, then absent contrary evidence, they should count as positive evidence. This is Alf and Betty's reasoning. Bundy does see the gap, as in his first essay, he identifies Alf as relying on a principle he terms 'suspension as evidence.'[4] (Bundy conceded the principle for the sake of the argument). The point is that this cluster of conditionals isn't simply PS. That's right, but note the tight connection between them. Again, PS and Alf and Betty's conditionals are all manifestations of the view that the opinions of peers function as evidence that can either defeat or further support. So Bundy's objection is correct – our case did not proceed exclusively from PS, but from a cluster of closely-tied commitments that are reflective of a broad class of views we'd identified as

[4] Bundy, "In Defense of Epistemic Abstemiousness," 290.

motivating what we'd called epistemic abstemiousness. The principles that yield our story, it is true, are not exclusively PS, but they are nevertheless tied together by a form of evidential parity – roughly that, if only evidence can determine our cognitive duties, then, if the beliefs of others can function as a defeater for one's justification, it must be evidence. PS and this parity principle provide norms Alf and Betty later follow. And so, yes, PS is not what yields Alf's distortion or Betty's conversion, *per se*, but the norms that do yield them are derived from it.

III

Bundy argues that Betty has reason to hold that Alf, because he does not follow PS, is not a peer. As Bundy puts it, "when Alf does not suspend judgment regarding p, [Betty] acquires [evidence that Alf retains his belief in the face of disagreement], which in turn gives her reason to think that Alf should not be fully trusted when it comes to p."[5] Betty then may "reasonably conclude that Alf is not a peer when it comes to p."[6]

The trouble with this line of thought, as we see it, is that the disagreement question in epistemology arose precisely because of the *persistence* of *deep* disagreements. The objective behind PS is to avoid being dogmatic in the face of these challenges, and so it seems positively strange to downgrade peerhood for others solely on the basis of their disagreement with PS.

Bundy responds that the apparent *strangeness* of this result is mitigated by two features of his view. Taking the second first: He insists that stubbornness on behalf another will not always count as reason for denying that person peerhood because sometimes such stubbornness will count as "evidence that the person is better positioned epistemically, and so is one's epistemic superior when it comes to evaluating whether p."[7] However, one could only reach that conclusion if one already had evidence that this person, *in addition to* his or her refusal to apply PS, was epistemically superior with respect to p (otherwise there would be no way to distinguish this person from the one whose failure to apply to PS undermines his or her status as a peer). But in that case, one is not dealing with an epistemic peer in the first place, and so the question of whether to apply PS does not even arise.

His more fundamental objection, the one that sheds most light on the deep difference in what we take to be at stake in arguments about PS, emphasizes that peerhood is "relative to a particular proposition." So one's determination that another is not a peer with respect to p is consistent with treating that person as

[5] Bundy, "On Epistemic Abstemiousness," 622.

[6] Bundy, "On Epistemic Abstemiousness," 622.

[7] Bundy, "On Epistemic Abstemiousness," 623.

otherwise smart, competent, well-informed, etc.[8] But our objection was never that Bundy's account was troubling because it would permit Betty to treat Alf as if he were generally an imbecile. Rather our concern was that this move is inconsistent with the spirit of PS which is supposed to be a principle for taking peer disagreement seriously. In the face of such disagreement, one should not remain dogmatic, but instead reconsider one's own deeply held beliefs. On Bundy's version of things, however, PS is a mechanism for *dismissing* certain cases of disagreement, namely disagreement with an apparent peer over whether it is appropriate to suspend judgment with respect to p. In such cases, the proper conclusion according to Bundy is not to revaluate one's own beliefs, but rather the other's peerhood. This strikes us as incongruent because we cannot think of any reason why PS should be special in this regard. That is: why should PS apply to all disagreement *except* disagreements over whether to apply PS? A consistent application of PS would prohibit the strategy Bundy advocates thus resulting in the descent into conversion and martyrdom we outlined in our original paper.

Bundy suggests that Adam Elga has answered just this objection with his argument for a "partially conciliatory view" which offers a principled method for taking disagreement about disagreement off the table of conciliation — that is, Elga argues that views on disagreement can be excluded as proper objects of PS. Elga notes, as we do, that this exclusion would be arbitrary without some independent motivation. He suggests that, "the real reason for constraining conciliatory views is not specific to disagreement. Rather, the real reason is a completely general constraint that applies to any fundamental policy, rule, or method. In order to be consistent, a fundamental policy, rule or method must be dogmatic with respect to its own correctness. This general constraint provides independent motivation for a view on disagreement to treat disagreement about disagreement in a special way."[9] First, this simply pushes the argument back a step. We still need to see an account of *why* disagreement norms generally, and specifically the PS, count as fundamental in this sense. It's clear that *if it did*, its exclusion from conciliation would be a non-arbitrary.

Even if such an argument is forthcoming, however, the cost of treating views of disagreement as fundamental is extremely high. The troubling upshot of Bundy's, as well as Elga's, view is that people who disagree about when to apply PS cannot be epistemic peers, or at least, we're justified in holding they aren't. In our original reply, we argued that this effectively renders disagreement amongst epistemic peers

[8] Bundy, "On Epistemic Abstemiousness," 623.

[9] Adam Elga, "How to Disagree About How to Disagree," in *Disagreement,* eds. Richard Feldman and Ted Warfield (New York: Oxford University Press, 2010), 185.

impossible: debates amongst such peers can never terminate in disagreement because either both parties will agree to suspend or they are not in fact peers. Bundy objects by noting that disagreement is still possible amongst epistemic peers who mutually fail to apply PS (that is, peers who disagree, but do not think that peer disagreement warrants suspension).[10] But surely *that* is an even stranger result. Recall that PS is supposed to urge us to take peer disagreement seriously, but now it turns out that the only people capable of acknowledging that they have epistemic peers with whom they disagree are those who *reject* PS, and thereby are *epistemic failures*! We have argued from the beginning that the appeal of PS, the appeal of the conciliatory view, is its promise to help in matters of fundamental, deep, disagreement between people who can reasonably regard each other in cognitively favorable lights. If PS rules this out in advance, it is not clear to us what the remaining appeal of the principle is. The only way out of this problem that we can see, is to take the diachronic dialectical-epistemic consequences of how we treat disagreement (that is, to treat just those consequences that made PS appealing in the first place) as relevant.

[10] Bundy, “On Epistemic Abstemiousness,” 624.

(MORE) SPRINGS OF MY DISCONTENT: A REPLY TO DOUGHERTY

Guy AXTELL

ABSTRACT: A further reply to Trent Dougherty, author of *Evidentialism and its Discontents*, on a range of issues regarding a proper understanding of epistemic normativity and doxastic responsibility. The relative importance of synchronic and diachronic concerns with epistemic agency is discussed, both with respect to epistemology 'proper,' as well as in connection with broader concerns with 'ethics of belief' and 'epistemology of disagreement.'

KEYWORDS: doxastic responsibility, synchronic and diachronic, ethics of belief, epistemology of disagreement

I want to thank Trent Dougherty for his comments in "Re-reducing Responsibility" on my paper in this journal, "Recovering Responsibility."[1] Dougherty raises concerns that further highlight key differences between the exclusively "synchronic" focus of his internalist evidentialism, and the more "diachronic" focus of virtue epistemologies generally, and of virtue responsibilisms (or character epistemologies) in particular.

Dougherty begins by providing examples of (standard non-epistemic) moral ir/responsibility and (standard non-epistemic) instrumental ir/rationality – forgetting to mail an important check, drinking too much, spending too much on a watch, and the like. He characterizes them by noting that "Neither of these has anything particularly epistemic about it," and I agree. He then goes on to reiterate his IT or *Identity Thesis* according to which "There are nothing but moral irresponsibility or practical irrationality in cases of epistemic irresponsibility" [or perhaps better, in cases that character epistemologists describe as illustrating epistemic irresponsibility].

[1] Trent Dougherty, "Re-reducing Responsibility: Reply to Axtell," *Logos & Episteme* II, 4 (2011): 625-632. Guy Axtell, "Recovering Responsibility," *Logos & Episteme* II, 3 (2011): 429-454. In that paper I critiqued Dougherty's "Reducing Responsibility: An Evidentialist Account of Epistemic Blame," a paper in which he was defending evidentialism as set forth by Earl Conee and Richard Feldman against criticisms. *European Journal of Philosophy*, 2011, doi: 10.1111/j.1468-0378.2010.00422.x.

My first response is simply to note again how un-intuitive is the claim IT makes. The "Craig Case" that our discussion has focused upon seems very dissimilar in basic respects to *any* of the author's cited 'pure' or 'standard' examples, since it directly concerns Craig's doxastic habits or dispositions – how he goes about maintaining confidence in the truth of his belief by flatly refusing to countenance or pursue counter-evidence brought to his attention. I would call this a question of Craig's doxastic ir/responsibility, but however we describe it, certainly there is *something* particularly epistemic about the case that must be recognized, *something* importantly disanalogous between it and those instances of 'pure' moral irresponsibility or practical irrationality he cites. I would think an evidentialist should grant this much, even if he wants to argue that these disanalogies aren't enough to lead us to treat inquiry or evidence-gathering activities (what responsibilists call *zetetic activities*) as a proper subject matter in epistemology.

Apparently, though, Dougherty, does not want to give up this much. He alleges that in treating as an epistemological concern Craig's blanket refusal to heed or read potential defeating evidence to his special creationist belief brought to his attention, I am inventing new, *sui generis* or 'emergent' sorts of normativity. We should instead be *reducing* the springs of normativity to their lowest number: one.[2] This charge isn't well-developed, aside from loose analogies, such as that "Being practically irrational with respect to some matter of belief does not result in some sui generis, emergent 'epistemic' irrationality any more than paying too much for a meal takes on some sui generis, emergent 'culinary irresponsibility.' The view that I am inventing a 'new' or 'sui generis' kind of normativity in speaking of evaluating activities of inquiry from an epistemic point of view, it should be noted, is an *interpretation* Dougherty provides of the consequences of my opting out of his intended reduction of epistemic normativity to evidential 'fit.' But who besides the evidentialist would have thought of the "epistemic" in such narrow terms? Historically, I do not think that epistemologists have treated issues of "doxastic responsibility" either as non-epistemological or as a purely synchronic matter of 'fit' with evidence regardless of how well or ill-gotten that evidence is. The position presented as a radical one on my part seems rather to have been germane to epistemological concern in the modern era, the heyday of Chisholmian internalism, roughly contiguous with that of logical positivism and its fact/value dichotomy

[2] His original paper also suggests plans to reduce moral irresponsibility to pragmatic irresponsibility in later works, leaving us with two source of normativity overall, the practical (including the moral) and the epistemic (which for Dougherty as for Feldman and Conee is wholly a matter of 'fit' between one's doxastic attitude and one's present evidence bearing upon a proposition).

being the main exception. I don't see this stance as having the consequence of multiplying kinds of normativity out of hand as Dougherty suggests, and my resistance of Dougherty's reductionist stance is, at any rate, consistent with my own developed "too narrow" objection to the evidentialist conception of epistemic normativity.[3]

One aspect of the reductionism that I argued against in my paper in Dougherty's recent collection, *Evidentialism and its Discontents*[4] is the claim made by Feldman that "By seeking out new evidence concerning some important proposition and then believing what the evidence supports, I don't do a better job of achieving the goal of believing reasonably. I achieve that goal at any moment by believing what is then supported by my evidence … the epistemically rational thing to do at any moment is to follow the evidence you have at that moment."[5] This is reductionist in the double sense that, firstly, the sources of epistemic value, which I view as plural, are reduced to the standard that evidentialists term *synchronically rational belief* (i.e. 'fit'); and, secondly, that "believing reasonably" is reduced to being "epistemically rational" (i.e., having doxastic attitudes that remain in constant 'fit' with what is taken as evidence, however gotten or ill-gotten.

To examine this, the first reduction imposes a particular account of how to "maximize epistemic value," an account that those not committed to internalist evidentialism reject. Responsibilists doubt that this standing – being synchronically rational – is of uniform epistemic value; its value is contingent upon a base level of expected doxastic responsibility. With questions of doxastic responsibility suspended, the type of rationality the evidentialist puts all their chips on appears to be of doubtful epistemic worth; it may even be a good way to get things wrong. On my view, Feldman's principle of epistemic value-maximization furthermore confuses the *contrastive* judgment that one ought to have the doxastic attitudes that fit one's evidence (rather than a doxastic attitude that doesn't), with the far stronger and more doubtful claim that epistemic *oughts* are exhausted by those that meet his principle. On any externalist account, including the "mixed" variety that

[3] Bernard Williams makes the related point that the more analytic philosophy in various areas has become, the more exclusively "synchronic" its focus has sometimes become. The responsibilists who Dougherty takes issue with in his paper would I think all agree that this a contingent historical circumstance. On my view virtue, social, feminist epistemologies, genealogical approaches, etc. represent a counter-trend that can provide needed balance by including developmental and longitudinal perspectives on epistemic agency as well as genealogical conceptions of the functions of our central epistemic concepts.

[4] Trent Dougherty, ed., *Evidentialism and its Discontents* (Oxford: Oxford University Press, 2011).

[5] Earl Conee and Richard Feldman, *Evidentialism: Essays in Epistemology* (Oxford: Oxford University Press, 2004).

responsibilists tend to favor, doxastic justification and doxastic responsibility cannot always be divided off from one another in the way Dougherty suggests. The causal origin or reliable etiology of a belief matters to its epistemic standing, and this in turn implicates the epistemic conscientiousness or sloth of the agent as an epistemological concern in certain 'problem cases' including the one at hand. In short, virtue reliabilists and responsibilists have both explicitly argued that the evidentialist account of account of epistemic value maximization is too narrow, and that this is a major weakness in the theory.[6]

But the second reduction in Feldman's passage is actually more interesting to me, in that it has dramatic, though rarely noticed practical consequences. This reduction of standards of *reasonable* belief to that of *synchronically rational belief* I would argue is a primary reason why the epistemologies of disagreement developed from this evidentialist starting point have been unable to support reasonable disagreement (or Rawlsian reasonable pluralism), and instead tend to devolve into a silly stalemate between supporters of a "Uniqueness View" ('if I'm rational, you're not') and an "Equal Weight View" ('whenever evidence-sharing peers disagree, suspension of judgment is our automatic epistemic duty').

[6] Virtue reliabilists and responsibilists essentially agree on these points about evidentialism leading to a denaturing of doxastic responsibility, and the diachronic aspects of epistemological evaluation more generally. When doxastic justification and thus "the knowing-self moves to center stage, epistemic evaluation, whether it is of beliefs or of character, cannot function within the constraints of a strict internalism. The relaxation of internalist criteria occurs on two fronts. First, consideration of reliability and success in achieving truth become relevant, and second, a social dimension is introduced to rupture the isolationism of purely 'internal' looks within' … Epistemic responsibility now is not a function of either not violating epistemic obligations (deontology) nor of factors purely transparent to the knower (internalism). Vrinda Dalmiya, "Knowing People," in *Knowledge, Truth, and Duty*, ed. Matthias Steup (Oxford University Press, 2001), 232. Alvin Goldman relatedly writes, "The main problem facing deontological evidentialism is to account for the virtues of evidence gathering. If proportioning your degree of belief to the weight of evidence is the sole basis of epistemic virtue, epistemic agents can exemplify all virtues without gathering any evidence at all, by working with the most minimal quantities of evidence … [I]t is just as meritorious for an agent to adopt a doxastic attitude of 'suspension' when her evidence is indecisive as it is for her to adopt a doxastic attitude of full conviction when her evidence is quite dispositive. No further epistemic merit or praise can be earned by investigation, research, or clever experimentation the outcome of which might discriminate between competing hypotheses. In short, deontological evidentialism is perfectly content with investigative sloth! This is surely a major weakness of the theory, because numerous epistemic virtues are to be found among processes of investigation." Alvin I. Goldman, *Pathways to Knowledge: Private and Public* (New York: Oxford University Press, 2002), 56.

Returning to our central argument, Dougherty doesn't take the obvious fact that doxastic responsibility is truth-directed, and that some strategies and habits are reliable and others unreliable in leading to true beliefs, as doing anything to qualify the inquiry-directed efforts of an agent as an epistemological concern. But I think he fails to see the force of the point that an agent who conducts him or herself as Craig does will manifest very well-known cognitive biases: We could hardly imagine that Craig, in refusing to countenance or pursue in any way the counter-evidence to his belief, wouldn't be committing what standard critical thinking textbooks refer to as "fallacies of relevance." I am speaking of fallacies like *Appeal to Consequences* ("I can't read or consider that recommended book on evolution because it will lead to ungodliness"); *Appeal to popularity* ("Others tell me not to read such rubbish, so rubbish it must be"). *Weak induction* and *improper appeal to authority* are other candidate fallacies of relevance that come to mind for the Craig case or similar cases like those that Baehr and DeRose discuss.[7] Dougherty and Feldman are committed to claiming that Craig is "doing fine" qua epistemic agent so long as he remains synchronically rational, but isn't there an obvious inconsistency here? How is it that it's at best a *moral* or *pragmatic* shortcoming to manifest *cognitive* biases, and to commit fallacies of relevance?

To even have the ability (wherewithal) to avoid fallacies of relevance one must be able to distinguish genuine evidence from various forms of emotional appeal, etc. There are skills at issue here – skills and habits. Dougherty thinks he has a reply this. He responds that "Failures due to lack of skill might be sad or comical but they can't be cases of any kind of irresponsibility as far as I can see unless the fact of the lack of skill came about via a moral or prudential shortcoming." I don't find this reply satisfactory. I'm firstly not saying it's as simple as that to manifest moral biases is *exclusively* a moral fault, and cognitive biases *exclusively* an intellectual one. My position, like that of most responsibilists, is one that insists upon more 'entanglement,' and that suggests a diachronic encroachment on the purportedly 'pure' epistemic sphere of evidence as the evidentialist understands it. So I don't see why we should be committed to viewing competence over a normal level of intellectual habits and skills as a purely pragmatic or moral concern. It is a concern with the agent's intellectual competence, though no doubt it can often be looked at in these other ways as well. I argued for what I think is the common-sensical view that depending on the case, the *zetetic activities* of agents – their inquiry-directed activities – are assessable in light of moral, pragmatic, or

[7] See the papers by Jason Baehr and Keith DeRose for further development and depth discussion of such problem cases, and Conee and Feldman's responses, in Dougherty, ed., *Evidentialism and its Discontents*.

epistemic (truth-directed) norms. The question depends on the case, and also on our interests in explanation, but the fact that it is "activities" and diachronic aspects that are in question does not necessarily push the issue outside of the purview of epistemological interest. This multiple-assessability model I developed contrasts sharply and I think quite advantageously with the presented view of Dougherty, in which everything is either a purely this or purely that, and in which all such activities of agents, being part of the active aspects of agency, can only be assessed morally or pragmatically.

My final comment, picking up from where we began, is on the interesting points about the 'ethics of belief' that Dougherty makes. Dougherty charges responsibilists generally and me in particular with a conflation: "By failing to realize that the ethics of belief is just a kind of applied ethics, serious mistakes are made about the nature of epistemic justification, knowledge, and other forms of positive epistemic status." Now the issues that Dougherty utilized the Craig case to raise were decidedly *not* those of the ethics of belief, or of what Craig has a 'right' to believe, all things considered. If we were speaking of the ethics of belief, ethics would be directly pertinent, and we would be debating the proper way to restrict the domain in which instances of believing may be judged on ethical as well as epistemic grounds. That has not been our focus by a long shot, so that I am somewhat surprised at Dougherty's remark. But I take it that what he means is that the diachronic concerns I and other responsibilists raise are *properly* relocated as concerns with an ethics of belief (in contrast with epistemic appraisal). Although I rarely see a point to the 'just' and 'nothing buttery' talk, I certainly do see a potential worry here. Certainly, first of all, the synchronic/diachronic divide is evident in alternative views about the norms that should inform an ethics of belief; certainly as well, it is possible to make the 'serious mistake' Dougherty alleges, in conflating these issues. Perhaps then evidentialists and responsibilists are *both* prone to conflating these issues, but in *opposite* ways. What I mean is that, if readers think that responsibilists bring too much of the diachronic into questions of epistemic appraisal, we should also think about the serious mistakes made by evidentialists like Feldman, who 'reduce' the ethics of belief and the question of the possibility of reasonable disagreement to a branch of applied epistemology. If, as I have elsewhere argued, the norms that should inform a sound and civic ethics of belief are primarily diachronic, what is 'out of place' is rather the primacy of synchronic rationality in Feldman's ethics of belief.[8] But rather than trying to

[8] For a critique of Feldman's ethics of belief and epistemology of disagreement, see my "From Internalist Evidentialism to Virtue Responsibilism," in *Evidentialism and its Discontents*, 69-87. For a positive development of a substantially more liberal account of the ethics of belief, an

defend myself further against Dougherty's charge, let me end by just asking: Can't we make equally serious mistakes about the nature of the norms that should inform the ethics of belief (and the understanding of peer disagreement) by failing to realize that internalist evidentialism is properly just an analysis of evidential justification?

account where backward and forward-looking diachronic responsibilities are foremost, and responsibility is clearly distinguished from the narrower norm of synchronic rationality, see my "Possibility and Permission: A neo-Jamesian Ethic of Belief," in *William James' Philosophy of Religion*, eds. Sami Pihlström and Henrik Rydenfelt (Basingstoke: Palgrave Macmillan, 2012, forthcoming).

A NOTE ON ASSERTION, RELATIVISM AND FUTURE CONTINGENTS

J. Adam CARTER

ABSTRACT: I argue that John MacFarlane's attempt to reconcile his proposed truth-relativist account of future contingents with a plausible account of assertion is self-defeating. Specifically, a paradoxical result of MacFarlane's view is that assertions of future contingents are impermissible for anyone who already accepts MacFarlane's own truth-relativist account of future contingents.

KEYWORDS: assertion, truth-relativism, future contingents

Do future contingents have truth values? This is an important question for the purposes of theorising about assertion, and in particular, assertoric norms. Norms of correctness govern assertions, and these norms are epistemic in nature.[1] For example: "assert p only if you know p"[2] or, more weakly, "assert p only if p is true."[3] If future contingents don't have truth values – if presently it is neither true nor false that "There will be a sea battle tomorrow" – then, if (for instance) either the knowledge or truth norm of assertion is correct, the assertion "There will be a sea battle tomorrow" is a defective assertion.[4] Moreover, if it is impermissible to assert "There will be a sea battle tomorrow" given one's epistemic grounds, then plausibly

[1] For an overview of recent work on norms of assertion, see Jennifer Lackey, "Norms of Assertion," *Noûs* 41 (2007): 594-626.

[2] e.g. Timothy Williamson, "Knowing and asserting," *The Philosophical Review* 105, 4 (1996): 489-523, *Knowledge and its Limits* (Oxford: Oxford University Press, 2000); Keith DeRose, "Assertion, Knowledge, and Context," *The Philosophical Review* 111 (2002): 167-203; John Hawthorne, *Knowledge and Lotteries* (Oxford: Oxford University Press, 2004); Jason Stanley, *Knowledge and Practical Interests* (Oxford: Oxford University Press, 2005).

[3] e.g. Michael Dummett, "Truth," *Proceedings of the Aristotelian Society* 59 (1959): 141-62; Matthew Weiner, "Must We Know What We Say?" *Philosophical Review* 114 (2005): 227-251. A middle ground 'justificationist account of assertion' has been defended recently by Douven (Igor Douven, "Assertion, Knowledge and Rational Credibility" *The Philosophical Review* 115 (2006): 449-485), Lackey (Lackey, "Norms of Assertion") and Kvanvig (Jonathan Kvanvig, "Assertion, Knowledge, and Lotteries," in *Williamson on Knowledge*, eds. Duncan Pritchard and Patrick Greenough (Oxford: Oxford University Press, 2009), 140-160). Roughly, the view is: assert p only if you are justified in believing that p is true.

[4] I take this example from John MacFarlane "Future Contingents and Relative Truth," *The Philosophical Quarterly* 53, 212 (2003): 321-336.

© LOGOS & EPISTEME, III, 1 (2012): 139-144

it is impermissible to use "There will be a sea battle tomorrow as a premise in one's practical reasoning."[5] But, providing one's evidence sufficiently supports a sea battle taking place tomorrow, it seems entirely permissible to assert "There will be a sea battle tomorrow" and it seems perfectly rational to use this as a premise in one's practical reasoning.

The natural response to these considerations is to suppose future contingents must have truth values; this is the determinacy intuition: an intuition that gains additional support from the thought that, when taking a retrospective view, utterances that 'turned out true' were true at the time of utterance[6]. Accordingly then, when I assert "There will be a sea battle tomorrow," my assertion counts as true if, tomorrow, there is a sea battle.

This result stands at odds with the indeterminacy intuition that, at the time of the utterance, multiple histories are possible, including one where there was a sea battle, and the proposition is true and one where there is not, and the proposition is false. The indeterminacy intuition leads us to think the truth value of future contingents is indeterminate at the time of utterance, and either true or false at a later time. John MacFarlane[7] thinks that both the indeterminacy intuition and the determinacy intuition should be taken at face value and that the only way to account for the semantics of future contingents is to allow the truth values of future contingents to be doubly relativised: to both the context of utterance and the context of assessment. On MacFarlane's proposal, when we evaluate the future contingent "There will be a sea battle tomorrow," this counts as neither true nor false when the context of assessment is the context in which the utterance is being made (as multiple possible histories are presumed open at this point). If the context of assessment is the following day, when there is a sea battle, the statement is 'true' and if there is not one, 'false.'

A key element of MacFarlane's position is that it rejects an assumption of the absoluteness of utterance-truth: the assumption that the truth value of an utterance

[5] It has become recently popular to suggest that knowledge is the epistemic norm of practical reasoning. For an especially clear presentation of this position, see Jessica Brown, "Knowledge and Practical Reason," *Philosophy Compass* 3, 6 (2008): 1135-1152. See, however, Mikkel Gerken, "Warrant and Action," *Synthese* 178, 3 (2011): 529-547, for a plausible case in favour of thinking that knowledge will be (many times) required to warrant action even though the matter of whether it, or merely justification, is required to warrant action shifts across contexts.

[6] As MacFarlane notes, it is commonplace to reason as follows: "Jake asserted yesterday that there would be a sea battle today / There is a sea battle today / So Jake's assertion was true." (MacFarlane, "Future Contingents," 325)

[7] MacFarlane "Future Contingents."

is independent of the context from which the utterance is being assessed.[8] In opposing the absoluteness of utterance-truth, MacFarlane's position on the semantics of future contingents is markedly relativistic.[9] The slippery slope from relativist semantics for future contingents to a more wide-ranging relativist semantics doesn't bother MacFarlane. "Future contingents are important because they force us to abandon absoluteness, liberating us from its conceptual bonds elsewhere."[10]

What I'm interested in engaging with here is not the big-picture worry regarding the implications of a relativist semantics for future contingents for other cases. My focus will be on the matter of whether MacFarlane's relativist semantics for future contingents is plausible. And on this score, my focus will be assertion. MacFarlane recognizes that rejecting the absoluteness of utterance-truth assumption stands in some tension with providing a plausible account of assertion. He attempts to reconcile this problem, but I do not think he does so successfully. MacFarlane's attempt to reconcile his relativism about future contingents with a plausible account of assertion stems in part from his attempt to reply to a potential objection from Gareth Evans[11] on this score. As Evans writes:

> Just as we use the terms 'good' and 'bad', 'obligatory' and 'permitted' to make an assessment, once and for all, of non-linguistic actions, so we use the term 'correct' to make a once-and-for-all assessment of speech acts .... if a theory of reference permits a subject to deduce merely that a particular utterance is now correct but later will be incorrect, it cannot assist the subject in deciding what to say, nor in interpreting the remarks of others. What should we aim at, or take others to be aiming at? Maximum correctness? But of course, if he knew an answer to this question, it would necessarily generate a once-and-for-all assessment of

[8] MacFarlane's preferred 'truth-relativism' (in several areas of discourse) holds the truth-values of utterances to be determined always in part by a context of assessment. As Crispin Wright puts it: vary [the context of assessment] and the truth value of the utterance can vary, even though the context of its making and the associated state of the world remain fixed (Crispin Wright, "New Age Relativism and Epistemic Possibility: The Question of Evidence," *Philosophical Issues* 17, 1 (2007): 262-283).

[9] MacFarlane has defended truth-relativism in various domains of discourse including epistemic modals, predicates of personal taste and knowledge attributions. For lucid presentations of MacFarlane's truth-relativism, see his "Making Sense of Relative Truth," in *Proceedings of the Aristotelian Society* 105 (2005): 321–39. Reprinted in *Relativism: A Compendium*, ed. Michael Krausz (New York: Columbia University Press, 2010). For a helpful outline of MacFarlane's faultless-disagreement-style argumentative strategy for defending truth-relativism in other areas, see his "Relativism and Disagreement," *Philosophical Studies* 132 (2007): 17-31.

[10] MacFarlane, "Future Contingents," 336.

[11] Gareth Evans, "Does Tense Logic Rest on a Mistake?" (1985) in Gareth Evans, *Collected Papers* (Oxford: Oxford Clarendon Press, 2005), 346-63.

> utterances, according to whether or not they meet whatever condition the answer gave.[12]

MacFarlane's reply to Evans is nuanced. He claims that in making an assertion, one commits oneself to the truth of the claim (and so MacFarlane recognizes something like the truth norm for assertion); however, the kind of commitment this is specifically is a commitment to produce a justification – that is "giving adequate reasons for thinking that the sentence is true (relative to its context of utterance and the asserter's current context of assessment), whenever that assertion is challenged."[13] Call this, following Teresa Marques[14] the "meet-the-challenge" norm. Applying this view: if someone challenges (today) MacFarlane's protagonist (Jake)'s assertion (yesterday) that "There will be a sea battle tomorrow," "Jake can meet the challenge by pointing to ships fighting."[15] This is fine and well. But the problem arises for MacFarlane with respect to the way his view handles Jake's utterance "There will be a sea battle tomorrow" when the context of assessment is the same as the context of utterance. Following MacFarlane, let m_0 be the point at which the utterance is made (and a sea battle will not have either occurred or failed to occur until tomorrow). Here MacFarlane says (with a bit of background):

> In asserting "There will be a sea battle tomorrow" at m_0, Jake comes to be bound by certain obligations. For example, if someone challenges the assertion at m_0, Jake must give adequate reasons for thinking it is true, relative to the context of utterance m_0 and context of assessment m_0. If the challenge takes the form of a conclusive demonstration that it is not yet settled whether there will be a sea battle, Jake will not be able to meet the challenge, and he will be obliged to withdraw his assertion. But if the challenge is weaker, and he meets it, his assertion can stand.[16]

The problem here is that MacFarlane's promissory note – that if the challenge is weaker, Jake's assertion can stand – is not one that can be upheld. In

[12] Evans, "Does Tense Logic," 349. Greenough has sought to encapsualte the key elements of Evans's challenge as follows: (1) The question 'What should [an assertor] aim at?" is a legitimate question. (2) Any legitimate answer to this question will generate a once-and-for-all answer. (3) Any once-and-for-all answer is incompatible with Truth-Relativism (4) Therefore, Truth Relativism is ruled out (Patrick Greenough, "Relativism, Assertion and Belief," in *Assertion*, eds. Jessica Brown and Herman Cappelen (Oxford: Oxford University Press, 2010), 2).

[13] MacFarlane, "Future Contingents," 335.

[14] Teresa Marques, "Relativism and the Norm of Assertion," *LANCOG – Seminar Series in Analytic Philosophy 2008-09*, http://www.lancog.com/sem0809.html (last visited February 15, 2012).

[15] MacFarlane, "Future Contingents," 335.

[16] MacFarlane, "Future Contingents," 335.

fact, a direct implication of MacFarlane's relativist view will be that Jake's assertion is never permissible. Generalizing from this, we get the *reductio* that no future contingent assertions are permissible (and, likewise, no future contingents can viable to use as premises in practical reasoning). Why does MacFarlane's promissory note not hold up? This is because, put roughly, an individual S cannot provide an adequate justification for believing some assertion φ when S is not justified in believing that φ is true. Let's revisit the case of Jake, who asserts (at m_0) "There will be a sea battle tomorrow." Relative to the context of assessment at m_0, Jake's statement is neither true nor false but indeterminate. This is the result MacFarlane wants. However, MacFarlane can't get this result as well as the result that Jake's assertion is not epistemically defective. Even if we grant MacFarlane's preferred epistemic norm governing assertion – a sort of justificationist[17] norm according to which the rule is "assert p only if you can adequately justify p [to a potential challenger] at the time of assertion" – Jake fails to be justified in believing what he asserts. Though that's not quite right: more precisely, Jake would fail to be justified in believing what he asserts *if he also accepts that MacFarlane's relativism about future contingents is correct*. For if Jake does accept MacFarlane's account of future contingents, then Jake would not be able to adequately justify that his assertion "There will be a sea battle tomorrow" is true given that he accepts implicitly that it is (on MacFarlane's semantics) not true, but rather, neither true, nor false. So even if the challenge to Jake at m_0 was, as MacFarlane intimates, a 'weak challenge,' Jake will not in principle be able to provide an adequate justification for believing what he asserts as true, given his implicit belief that it is not (at m_0) true.

[17] I am taking it that MacFarlane's variety of a justificationist norm of assertion is a close cousin of the sort of justificationist norm of assertion defended by Lackey (Lackey, "Norms of Assertion"), Douven (Douven, "Assertion, Knowledge") and Kvanvig (Kvanvig, "Assertion, Knowledge"). Where MacFarlane's account comes apart from these other justificationist views is that the traditional justificationist position articulates the epistemic norm as one satisfied just in case one possesses certain reasons or evidence for the asserted proposition. MacFarlane on the other hand advances the specific requirement that one be able to provide such a successful justification to a challenger. A case where these two accounts come apart will be one where the agent's justification, though successful in response to a challenge, is not itself one the agent is justified in believing. For example, suppose I justify to a challenger my intentionally deceptive assertion that "The house is not for sale" to a challenger by pointing to a yard with no for-sale sign, even though I know that that the house has been put on the market that day and that the sign will be put up tomorrow. I take it that my case (awkwardly) satisfies MacFarlane's variety of the justificationist norm while violating the more traditional version of a justificationist norm, according to which the assertion would be epistemically defective. That said, MacFarlane's version will nonetheless align with the traditional version is a wide variety of cases.

So MacFarlane gets the awkward result that Jake is permitted to assert what he does only if Jake doesn't already accept MacFarlane's theory. This result is simply unacceptable. MacFarlane might reply by saying that Jake's justification of his belief (when challenged) at m_0 is successful so long as he justifies why it would be permissible to act as if his assertion were true. But to go this route would be to give up entirely on the view that assertions, as a category of speech act, are governed by any properly epistemic norm, and this would be an equally problematic result. I think the considerations given here are a serious mark against a relativist semantics for future contingents.

STILL NO SUICIDE FOR PRESENTISTS: WHY HALES' RESPONSE FAILS[1]

Jimmy Alfonso LICON

ABSTRACT: In this paper, I defend my original objection to Hales' suicide machine argument against Hales' response. I argue Hales' criticisms are either misplaced or underestimate the strength of my objection; if the constraints of the original objection are respected, my original objection blocks Hales' reply. To be thorough, I restate an improved version of the objection to the suicide machine argument. I conclude that Hales fails to motivate a reasonable worry as to the supposed suicidal nature of presentist time travel.

KEYWORDS: presentism, Steven Hales, suicide machine argument, time travel

Presentists hold everything that exists must occupy the present moment. The present moment exists to the exclusion of all other moments; this is because whatever would be located in the past/future does not exist. On the other hand, eternalists hold that time is similar to space in that all moments exist; one particular moment does not exist to the exclusion of any other moment. The present moment only appears special from the epistemic perspective of a specific occupant, just like locations seem to be privileged from some particular perspective.

Hales' suicide machine argument[2] holds that one cannot time travel in a presentist universe. If the only moment that exists is the present moment, and time travel amounts to leaving the present moment, one could not time travel; to do so would be tantamount to suicide as the time traveler would have to leave the present moment (all of reality).

In my previous response to Hales, I argued the suicide machine argument fails.[3] Although the presentist holds the present moment exists to exclusion of all other moments, presentism itself does not block the possibility of changing the structure of the present moment. If there is a machine which could re-arrange all of the matter and energy such that it was identical with a time other than the present, then time travel would be possible in a presentist universe.

[1] I would like to thank Professor Hales for his professional and thoughtful response to my work.

[2] Steven D. Hales, "No Time Travel for Presentists," *Logos & Episteme* I, 2 (2010): 353-360.

[3] Jimmy Alfonso Licon, "No Suicide for Presentists: A Response to Hales," *Logos & Episteme* II, 3 (2011): 455-464.

Hales holds his argument survives my objection.[4] His response rests on two key claims:

> a) The supposed time machine featured in my response is not up to the task of time travel as it merely rearranges all of the matter/energy in the universe, instead of transporting the time traveler to a different moment.
>
> b) The assumption of Humean supervenience [i.e., two moments of time are identical just in case they have the same arrangement of energy and matter] prevents someone from going back in time unless they were part of what constituted a past/future moment; but such a moment would not include a time machine/traveler.

I respond to both claims. In the last section, I argue that even if my objection to the suicide machine fails, Hales has not made his case that presentist time travel is suicidal.

A Better Time Machine

In my original article, I argued that machine F is a time machine only if (a) F is capable of rearranging all of the matter/energy in the universe such that it is indistinguishable from a moment in the past/future, and (b) the identity of indiscernibles holds between the instantiated and the past/future moment.[5] However, Hales thinks this conception of a time machine is inadequate.

For example, Hales writes:

> Suppose we set the controls of Licon's time machine for one month into the future. According to Licon, this means that the entire universe undergoes a somewhat radical reconfiguration of matter/energy. *Yet why should we consider this new state of the universe 'one month in the future'?* There was no alternative future history of the universe, no other calendar on which we can show the days that were skipped *or sped through by the time machine*. The sole history of the universe involves an unusual redistribution of matter/energy at a certain point, but that doesn't mean that anything traveled in time, or jumped one month into the future. The universe was in state A at one moment and in state B the next moment.[6]

In this passage, there appear to be two objections to my conception of a time machine.

First, Hales holds that *merely* rearranging all of the matter/energy of the universe such that it resembles a past/future moment does not constitute traveling

[4] Steven D. Hales, "Reply to Licon on Time Travel," *Logos and Episteme* II, 4 (2011): 633-636.

[5] Licon, "No Suicide," 462.

[6] Hales, "Reply to Licon," 634-4 (emphasis mine).

through time. It is not enough just to reproduce a moment that resembles a past/ future moment; there must be something else that makes such a process a kind of time travel. If a machine is capable of rearranging all of the matter/energy from the universe, there must be a difference-maker between scenarios where (a) machines merely create a moment that resembles a past/future moment without that operation constituting time travel and (b) machines that re-arranging all of the matter/energy in the universe, along with some other factor, such that it constitutes time travel (e.g. re-creating all of the moments that connect the departure and arrival moments).

Second, Hales argues the conception of a time machine featured in my objection ignores the distinction between personal and external time. If a presentist time machine is designed solely to rearrange all of the matter/energy in the universe to the destination time for the traveler, then there can be no distinction between time from the perspective of the time traveler (personal) and time from the perspective of everyone else (external). Traveling from the early twenty-first century to the Middle Ages from the perspective of the time traveler and a calendar external to the time machine, time would instantaneously change from the twenty-first century to the Middle Ages. However there should be a distinction between the personal time of the time traveler and external calendar time. Although traveling from the present to the past would be instantaneous from the perspective of the time travel, it would be a longer time in time on the calendar; e.g. the time separating the present moment from the Middle Ages is far more than a few seconds.

Unfortunately, the time machine in Hales response only mildly resembles the time machine featured in my objection as it neglects a key component of my time machine proposal: Leibniz's law of the identity of indiscernibility[7] must hold between the instantiated moment and a past/future moment. Hales is quite right that a machine which merely rearranging all of the matter/energy in the universe to resemble a moment from the past/future does not qualify as a time machine. If a machine rearranges the universe such that the instantiated and past/future moments respect the identity of indiscernibles, and the instantiated moment exists to the exclusion of all other moment, then it is a plausible candidate for a presentist time machine.

The same response applies to Hales' claim that my conception of a time machine blurs the difference between personal/external times. Although the version of a time machine offered in my response to Hales' just rearranges all of the matter/energy of one moment such that it is arranged other-moment-wise, it

[7] *Leibniz's law of the identity of indiscernibility:* If, for every property F, object x has F if and only if object y has F, then x is identical to y.

nevertheless distinguishes between personal/external times. Although the time traveler moves from the departure to arrival moment instantaneously, the arrival/ departure times are separated by the moments in between in terms of their temporal ordering; it is a fact of the matter that years 1890 (arrival) and 2012 (departure) are separated by a significant amount of time, even if the time between the arrival/departure moments is never actually instantiated.

If Hales finds this clarification dissatisfying, there is a simple solution: add a constraint to the conception of the presentist-friendly time machine. For example, in addition to rearranging all of the matter/energy in the universe such that it (a) resembles a moment from the past or future, and (b) such that the identity of indiscernibles holds between the instantiated moment and a moment from the past/future, a presentist-friendly time machine must take such steps (a) and (b) to recreate every moment that connects departure/arrival moments.[8]

Humean, Schmumean

Next, Hales argues that if my response to his argument rests on the Humean supervenience conception of time, then my response fails. The Humean supervenience conception of temporal identity holds that moments are identical just in case their arrangement of matter/energy is identical. Hales argues the Humean assumption creates a problem for the possibility of time travel: if a past/future moment did not already contain the time traveler, then they could not travel to that particular moment. This is because a machine would create a past/future moment missing the relevant time machine/traveler.

Hales writes:

> If all of the matter/energy in the present moment is instantly rearranged to the exact configuration of matter/energy in 1862, then no one traveled back in time. '1862' is a rigid designator denoting a particular arrangement of matter and energy, and Licon's Humean supervenience constraint entails *that 1862 is recreated down to the smallest detail*. A recreation of 1862 does not allow for some matter to be differently assembled so that it forms a 'time traveler' from the future on the grounds that such an arrangement would not be 1862.[9]

Hales holds that Humean supervenience commits one to the following:

[8] Hales states the following: "… Yet why should we consider this new state of the universe 'one month in the future'? There was no alternative future history of the universe, no other calendar on which we can show the days that were skipped or *sped through* by the time machine" (Hales, "Reply to Licon," 634-5, emphasis mine).

[9] Hales, "Reply to Licon," 636 (emphasis mine).

A. If an instantiated moment is identical to a past/future moment, then the matter/energy of the instantiated moment must be arranged exactly like the matter/energy of the past/future moment.

But it does not follow from (A) alone that a past or future moment would not contain a time traveler. For that, Hales needs an additional proposition:

B. The past/future moment does not contain a time traveler.

The presentist lacks the motivation to accept (B). If the past/future moment was already constituted, in part, by the time traveler, i.e. it was a fact that someone successfully traveled to a past/future moment,[10] then an instantiated moment is identical to a past/future moment *only if* it was partly constituted by the time traveler. If someone traveled to a moment in the past, and a time machine instantiated a moment from the past/future, then that moment *should* contain the time traveler; otherwise, the moment differs from the past/future moment; the presentist should accept (B) only if she was convinced that time travel is not possible.

Furthermore, Hales thinks that if my response assumes a Humean supervenience view of time, then one moment is identical to another just in case the arrangement of matter/energy is identically arranged. Thus, a machine could temporally transport a time traveler to another time only if the arrival moment was already constituted, in part, by the time traveler. Hence, if Humean supervenience holds, one of following possibilities must be the case:

i. A time machine cannot rearrange matter/energy such that it meets the indiscernibility of identity for some past/future moment as this would not be time travel.[11]

ii. The time machine and its contents (e.g. the time traveler) do not constitute any part the past/future moment.[12]

iii. If someone travels to a past/future moment, then it is a matter of fact that they are part of what constitutes that past/future moment.

[10] Someone might wonder if there could be a truthmaker for such a proposition. I assume for the sake of this paper that there is a solution to the truthmaker problem for presentism as this paper is concerned with the possibility of presentist time travel. For a potential response to the truthmaker objection to presentism: Alex Baia, "Presentism and the grounding of truth," *Philosophical Studies*, forthcoming.

[11] A moment is identical to another moment if they have the same arrangement of matter/energy. Thus, if the past/future moment did not contain a time machine/traveler, the instantiated moment would violate the identity of indiscernibility: Hales, "Reply to Licon," 636.

[12] It might be that time travelers cannot alter past/future to which they traveled because they are not part of what constitutes the past/future. For example: Nicholas J. J. Smith, "Bananas enough for time travel?"*British Journal for the Philosophy of Science* 48, 3 (1997): 363-389.

Option (i) would undermine my response to Hales. However, unless Hales can motivate the inadequacy of options (ii) and (iii), the Humean supervenience assumption[13] is not a problem for my response to the suicide machine argument. Hales has failed to explain why a machine that respects the identity of indiscernibility, with the capacity to rearrange moments such that they are indistinguishable from past/future moments, that respects the distinction between personal and external time, and the exclusive conception of the present is not a time machine.

For example, consider the Sally thought experiment:

> Suppose that in a presentist universe, Sally enters a time machine, twirls the knobs to a time in the past and activates the machine. The time machine then proceeds to instantiate each moment between the departure moment and arrival moment (all of which meet the indiscernibility of identity), each to the exclusion of all other moments. Sally eventually arrives at her destination, and exits the time machine.

The Sally thought experiment meets Hales' criticisms, i.e. there is a distinction between personal/external times, the arrival moment exists to the exclusion of all other moments and so forth. This scenario is consistent with the constraints of presentism, but would not result in Sally's annihilation. If the Sally thought experiment is consistent with presentism and coherent, then the suicide machine argument fails, unless it is significantly modified.

Conclusion

Hales claims the time machine I proposed was not actually a time machine as it collapsed the distinction between personal and external time and failed to provide a sufficient difference-maker between traveling through time and merely rearranging the matter/energy in the universe. I argued this criticism ignores a central component of the time machine I proposed: the identity of indiscernibles must hold between the instantiated moment and some past/future moment. If the identity of indiscernibles is respected, there is an objective distinction between personal/external

[13] The presentist can do little but accept Humean supervenience identity conditions of moments. In a presentist universe, what else could serve as identity conditions for moments other than that moment being arranged F-moment-wise? Suppose Bob traveled back to 1890 in a presentist universe. What makes the moment occupied by Bob identical to 1890? It would seem that the only answer is that the moment is arranged 1890-wise. It cannot be a relation between different moments because presentism denies that there are any other moments that could stand in such a relationship. If Humean supervenience is problematic, this reflects on presentism, rather than presentist time travel.

times, i.e. the time traveler instantly traveled, though as a matter of fact that the calendar changed substantially.

Finally, Hales' claims Humean supervenience prevents a time machine from traveling to a past or future moment, as such moments are identical to their arrangement of matter/energy and do not already contain a time machine or traveler. However, this only works if Hales has established that presentist time travel is suicide. Otherwise, it seems that the presentist could respond that if time travel in presentist universe succeeded, and if Humean supervenience holds of times, then it must be that the time traveler constitutes part of a moment in the past/future. Otherwise, Hales is correct: there could be no time travel to such a moment. Thus, Hales' fails to adequately respond to my objection to his suicide machine argument.

ARE REASONS EVIDENCE OF OUGHTS?

Franck LIHOREAU

ABSTRACT: In a series of recent papers Stephen Kearns and Daniel Star argue that normative reasons to ϕ simply are evidence that one ought to ϕ, and suggest that "evidence" in this context is best understood in standard Bayesian terms. I contest this suggestion.

KEYWORDS: reasons, evidence, oughts, normativity, rationality

Reasons as Positively Relevant

What is it for a reason to have a certain strength? And what is it to weigh a reason against another? According to Stephen Kearns and Daniel Star,[1] if we accept the following claim about normative reasons for action:

> (RA) Necessarily, a fact F is a reason for an agent A to ϕ iff F is evidence that A ought to ϕ (where ϕ is an action)[2]

an answer to those questions can be fleshed out in terms of a familiar concept of evidence on which we already have a good, independent grasp, namely the concept of "incremental" evidence understood in standard Bayesian terms (or "positive relevance"):

> (IE) E is evidence for H iff $Pr(H \mid E) > Pr(H)$

that is, iff the probability of H when E is added (to one's prior background information, as encapsulated by Pr) is strictly greater than the probability of H (on one's prior information) alone. By combining this definition of evidence with the general claim about reasons in (RA), we get straightforward answers to the questions we started with: the strength of a reason to ϕ is the degree to which it

[1] Daniel Kearns and Kenneth Star, "Reasons: Explanations or Evidence?" *Ethics* 119 (2008): 31-56; "Reasons as Evidence," in *Oxford Studies in Metaethics 4*, ed. R. Shafer-Landau (Oxford: Oxford University Press, 2009): 215-242; "Weighing Reasons," forthcoming in *Journal of Moral Philosophy*; "Reasons, Facts-about-Evidence, and Indirect Evidence," forthcoming in *Analytic Philosophy*.

[2] Kearns and Star, "Reasons as Evidence," 216.

raises the probability that one ought to ϕ, and the stronger the reason to ϕ, the more probable it is that one ought to ϕ.[3]

My purpose in this note is not so much to contest the otherwise intriguing "reasons as evidence" thesis defended by Kearns and Star as it is to cast doubt on the suggested appropriateness of a standard Bayesian understanding of evidence for making sense of the "strength" and the "weighing" of reasons for action. To this end, I offer two counterexamples to the idea that all reasons to ϕ increase the epistemic probability that one ought to ϕ,[4] thereby establishing that within the scope of the account of incremental evidence in (IE), the account of normative reasons for action in (RA) is too narrow: not all reasons to ϕ are evidence that one ought to ϕ. As we shall see in due course, this result carries over to other standard Bayesian accounts of evidence as well.

Evidentially Irrelevant Reasons

A reason can fail to raise the (epistemic) probability, and therefore, assuming (IE), to be evidence that one ought to do something – or so shall I argue. This situation can arise in two different ways.

First, when the reason simply is "evidentially irrelevant" to the corresponding ought: A fact *F* can be a reason to ϕ even when *F* does not affect – neither raises nor lowers – the probability that one ought to ϕ.

To see this, consider the following case (directly inspired by, albeit freely adapted from an example by Peter Achinstein[5]).

Suppose you enjoy drinking a certain soda so much that you usually buy it by batches of 100 bottles; and today, you drank one and only one bottle of that soda from such a batch – call it batch *b* –, and no one else did. Now, consider the following claims about *b*:

> (*E1*) Newspaper 1 reports that 99 out of the 100 bottles in b are contaminated by an extremely dangerous and highly contagious virus.
>
> (*E2*) Newspaper 2 makes the same announcement as Newspaper 1 about *b*.
>
> (*H*) You have drunk from a contaminated bottle.

[3] Kearns and Star, "Reasons as Evidence," 232. See also Kearns and Star, "Weighing Reasons."

[4] In contesting this idea, I side with John Brunero ("Reasons and Evidence One Ought," *Ethics* 119 (2009): 538-545). But the lesson I draw differs from his in scope and is somewhat less categorical. For a discussion of Brunero's arguments, see Kearns and Star, "Weighing Reasons."

[5] Peter Achinstein, "A Challenge to Positive Relevance Theorists: Reply to Roush," *Philosophy of Science* 71 (2004): 521-524.

Clearly, E2 is not evidence for H since it neither increases nor decreases the probability of H:

$$Pr(H \mid E2 \,\&\, E1) = Pr(H \mid E1) = .99.$$

Because (the contents of) the reports are the same, adding Newspaper 2's report does not make it more probable that you've drunk from a contaminated bottle than does Newspaper 1's report alone.

Now, as a matter of public health, your drinking from a bottle contaminated by an extremely dangerous and contagious virus creates an obligation for you to put yourself into quarantine and stay at home; and although there might arguably be exceptions, the probability that you ought to do so remains nonetheless a strictly increasing function of the probability of *H*. Therefore, since *E2* does not affect the probability of *H*, it does not affect the probability that you ought to put yourself into quarantine either. So, if (IE) is true, *E2* is not evidence that you ought to put yourself into quarantine (and is even evidentially irrelevant to such an obligation).

However, it should be clear and uncontroversial that the .99 probability of *H* is more than enough, given *E1* (and the relevant public health obligations), for *E2* to be a reason – and a good one – for you to keep yourself in quarantine. So, the probability that you ought to keep yourself in quarantine is not affected by *E2*, despite the fact that, given *E1*, *E2* is a reason for you to keep yourself in quarantine.

Now, it is not strained to think that it is not properly speaking the newspaper report, but rather your drinking from a possibly contaminated bottle, that is a reason here. But Kearns and Star cannot afford this thought. For they explicitly state that a report or announcement of a fact – e.g. a newspaper report to the effect that people are starving in Africa – "has just as good a case for being a reason [viz. to send money to Oxfam] as do more paradigmatic reasons (such as that people are starving in Africa)."[6]

Moreover, note that mentioning the possible unreliability of the newspapers would be irrelevant here. Whether we assume the reports to be truthful or not, the result is the same. Let *N1* be that Newspaper 1 tells the truth about batch *b* and virus *v*, and *N2* that Newspaper 2 also tells the truth about *b* and *v*. Since as a matter of fact $Pr(H \mid E2 \,\&\, N2 \,\&\, E1 \,\&\, N1) = Pr(H \mid N2 \,\&\, E1 \,\&\, N1)$, *E2* fails again to raise the probability of *H*, and therefore of your obligation to put yourself into quarantine, despite the fact that *E2* still remains a reason for you to put yourself into quarantine (given *N2* & *E1* & *N1* this time). Therefore, the question of the newspapers' reliability does not arise here.

[6] Kearns and Star, "Reasons, Facts-about-Evidence, and Indirect Evidence," msp 1; see also "Reasons as Evidence," 233-234.

So, it seems that a fact can be a reason for one to do an act even if, because it does not affect the probability that one ought to do this act, the fact is not evidence that one ought to do the act. A consequence would be that assuming (IE) as an account of evidence, the analysis of reasons in (RA) is too narrow.

An easy fix would be to suggest replacing > with the weaker ≥ in the "naïve" analysis of incremental evidence we started with:

(IE*) E is evidence for H iff $Pr(H \mid E) \geq Pr(H)$.

This suggestion – at odds with standard Bayesian approaches to evidence – would fix the problem in the case at hand, since the condition on the right-hand side would *ipso facto* be satisfied. But a more serious problem is lurking.

Negatively Relevant Reasons

Failure of a reason to raise the (epistemic) probability that one ought to do something can indeed stem from another source as well, namely, from the reason being "negatively relevant" to the corresponding ought. In other words, a fact F can be a reason to ϕ even when, instead of raising it, F lowers the probability that one ought to ϕ.

To see this, take the following example. You own a restaurant that serves exotic food, and some highly perishable good, g, is being shipped to you as part of a bigger batch of perishable goods. Consider the following claims about g:

(H) g has gone off.

(E1) the shipping takes n days.

(E2) g is shipped as part of batch b.

And suppose you have somehow determined that good g has a 90% chance of having gone off if the shipping takes n days:

(1) $Pr(H \mid E1) = .9$,

that 75% of the goods in batch b have already gone off by the time they are sent:

(2) $Pr(H \mid E2) = .75$,

that the shipping of good g as part of batch b has a very low .088 probability of taking n days:

(3) $Pr(E2 \,\&\, E1) = .088$,

and that there is a not much higher .075 probability that good g has gone off after a n-day shipping inside batch b:

(4) $Pr(H \,\&\, E2 \,\&\, E1) = .075$.

Then, by the definition of conditional probability, we get an 85% chance that good *g* has gone off if it was shipped in *n* days inside batch *b*:

(5) $Pr(H / E2 \,\&\, E1) = Pr(H \,\&\, E2 \,\&\, E1) / Pr(E2 \,\&\, E1) \cong .85$.

As a consequence, the probability of *H* is actually lowered by *E2*, since:

(6) $Pr(H \mid E2 \,\&\, E1) \cong .85 < Pr(H \mid E1) = .9$.

Now, as a matter of public health and food regulations, the circumstances in which a restaurant might be allowed to serve spoilt food to its customers are presumably very, very few. I believe Kearns and Star will see no objection in conceding that conditional normative principles exist whereby if something like the nonnormative fact that the food is spoilt obtains, then so does something like the unconditional normative fact that one ought to throw it away and not serve it.[7]

And the probability that you ought to throw g away and not serve it to your customers is thus presumably a strictly increasing function of the probability of *H*. So, the probability that you ought to throw *g* away too is presumably lowered by *E2*. So, if (IE) is true, *E2* is not evidence that you ought to throw *g* away (and is even evidence that you ought not to throw *g* away).

But *E2* is a reason to throw *g* away. This point is relatively unproblematic: not only does the fact that *g* was part of *b* constitute a reason to throw it away, it constitutes a *very good reason* to do so by most people's standards given that 75% of the goods coming from batch *b* have already gone off by the time they are sent. So, the probability that you ought to throw *g* away is lowered by *E2*, despite the fact that, given *E1*, *E2* is a reason for you to throw *g* away.

Therefore, a fact can be a reason for one to do an act even if, because it lowers the probability that one ought to do this act, the fact is not evidence that one ought to do the act. As a consequence, the left-to-right reading of (RA) is false within the scope of (IE) – no need to say that it is false within the scope of (IE*) as well.

Discussion

It goes without saying that our proposed counterexamples to (RA) are not isolated examples, that it is not difficult to generate myriads of structurally similar examples, and that many other probability assignments could have been used to reach the same conclusion that some reasons are not evidence for oughts. Also, if such examples constitute genuine cases of practical reasons, these cases are "standard" in that "they are examples where the relevant facts are transparent to the agent, that is, where

[7] cf. Kearns and Star, "Reasons as Evidence," 229.

there are no false beliefs playing any role in deliberation and there is no misleading evidence around clouding the water."[8]

Still, one could respond to our counterexamples in one of two ways.

The first is simply to ignore the intuitive pull we undeniably feel towards considering them genuine cases of practical reasons at all and somehow insist that there are not. However, this line of response is not available to Kearns and Star since those cases satisfy the various sufficient conditions they state for being cases of practical reasons. They explicitly defend that if a fact *F* "can play an appropriate role in one's reliably concluding that one ought to ϕ",[9] or if it "can play an appropriate public role in rationally convincing [someone] that she ought to ϕ and in rationally convincing other people that she ought to ϕ",[10] or else if "it is normally the case that if a fact relevanty similar to *F* obtains, then one ought to do something relevantly similar to ϕ-ing",[11] then that fact *F* is a reason to ϕ. But it is uncontroversial that a newspaper report on a contaminated batch of goods as in the first of our cases, or a product being part of such-and-such batch of merchandise as in the second case, typically constitute information that can help us determine what we ought to do, help us justify what we do, convince others about what they ought to do, and enter into the formulation of normative principles connecting relevantly similar information with obligations towards relevantly similar actions (as reflected in health conventions and food regulations, for instance). So, Kearns and Star will have to concede that the relevant facts involved in our putative counterexamples are indeed reasons to do the relevant acts.

The second way one could respond to our cases is to opt for a different Bayesian account of evidence. Among some of the other relatively standard options available, one is to drop the incremental notion of evidence in (IE) in favor of an "absolute" one:

(AE) E is evidence for H iff $Pr(H \mid E) > k$, for some degree k of high probability

while another option is to go for a "probative" notion of evidence instead:

(PE) E is evidence for H iff $Pr(H \mid E) > Pr(H \mid \neg\text{E})$.

Unfortunately, none of these options will work.

In (AE) the appropriate threshold k can undoubtedly be set very high. Still, in the context of the *reasons as evidence* thesis, it will have to be set low enough to

[8] Kearns and Star, "Reasons as Evidence," 223.

[9] Kearns and Star, "Reasons as Evidence," 225.

[10] Kearns and Star, "Reasons as Evidence," 227.

[11] Kearns and Star, "Reasons as Evidence," 228.

do justice to the fact that, by most people's standards, the facts involved in our purported counterexamples do count as reasons, and even good reasons to do the relevant acts: an 85% conditional probability that the food has gone off and a 99% conditional chance of having drunk from a contaminated bottle are, in this respect, more than high enough. So, the proposed examples are counterexamples to (RA) within the scope of (AE) too.

As to (PE), in the first of our cases the epistemic probability of having drunk from a contaminated bottle when Newspaper 2's report is added is not higher than the epistemic probability of having done so in the absence of Newspaper 2's report, since this probability already is 99% given the report made by the other newspaper, Newspaper 1. In our second case the epistemic probability that the food is not part of the incriminated batch *b* and takes *n* days to be shipped (i.e. ¬*E2* & *E1*) can easily be specified so that the epistemic probability that the food has gone off if it was part of this batch and took that long to be shipped is lower than the epistemic probability that it has gone off if it was not part of that batch yet took that long to be shipped. So, the objection that stems from our counterexamples carries over from (IE) to (AE) and (PE).

Conclusion

To sum up, the probabilistic implementation that Kearns and Star suggest to put flesh on the bones of their general *reasons as evidence* thesis is too narrow: (RA) fails to provide a necessary condition for being a normative reason for action within the scope of the Bayesian account of incremental evidence they suggest, (IE), and this is true as well with other standard Bayesian understandings of evidence, like the account of absolute evidence in (AE) and the analysis of probative evidence in (PE).

It would certainly not be fair to conclude from this to the inadequacy of (RA) itself. As Kearns and Star remark, the claim that results from combining (RA) with (IE) is more specific than the claim in (RA), and likewise with (AE) and (PE). But they do place hopes in the possibility of explaining what is involved in weighing reasons against each other in terms of a particular account of weighing evidence along standard Bayesian lines.

So what we may conclude from the foregoing considerations is that appeal to such standard Bayesian accounts of evidence as those in (IE), (AE), or (PE) seems inappropriate to Kearns and Star's general purpose of making sense of the notion of a reason's strength and of weighing reasons. Their hopes in this respect might not be so well-placed as they think.[12]

[12] An earlier version of this material was presented at the New University of Lisbon. I wish to thank the audience for many helpful comments. My work on this paper was carried out in part within the "Argumentation, Context, and Communication" project (PTDC/FIL-FIL 110117/2009) supported by the Portuguese Foundation for Science and Technology.

NOTES ON THE CONTRIBUTORS

Scott F. Aikin is Senior Lecturer at the Philosophy Department of Vanderbilt University. His areas of interest are: epistemology, informal logic, argumentation, and pragmatism. His books include: *Epistemology and the Regress Problem* (2011), *Reasonable Atheism* (2011) and *Pragmatism: A Guide for the Perplexed* (2008) – both with Robert B. Talisse. In *Logos & Episteme* he published, with Michael Harbour, Jonathan Neufeld, and Robert B. Talisse, "On Epistemic Abstemiousness: a Reply to Bundy" (2011) and "Epistemic Martyrs, Epistemic Abstainers, and Epistemic Converts" (2010). Contact: scott.f.aikin@vanderbilt.edu.

Guy Axtell is Assistant Professor in the Philosophy and Religious Studies Department, Radford University. His main teaching interests are in epistemology and metaphysics, philosophy of science, "STS" or science, technology and society studies, and philosophy of religion. His most recent publications include: "Recovering Responsibility" (*Logos & Episteme*, 2011), "Three Independent Factors in Epistemology" (with Phillip Olson, *Contemporary Pragmatism*, 2010), "Character-Trait Ascription in Ethics and Epistemology" (*Metaphilosophy* special edition, *Virtue and Vice, Moral and Epistemic*, 2010), and "Virtue Theoretic Responses to Skepticism" (in *Oxford Handbook of Epistemology,* ed. John Greco, 2009). Contact: gsaxtell@radford.edu.

Peter Baumann is Professor of Philosophy at Swarthmore College. His main areas of specialization are epistemology and philosophy of mind. He is also interested in questions concerning the nature and limits of rationality. He has published articles and monographs in these areas among which: "Involvement and Detachment: A Paradox of Practical Reason" (*Practical Conflicts*, eds. Peter Baumann and Monika Betzler, 2004), "Three Doors, Two Players, and Single Case Probabilities" (*American Philosophical Quarterly*, 2005), "Theory Choice and the Intransitivity of *Is A Better Theory Than"* (*Philosophy of Science,* 2005), "Is Knowledge Safe?" (*American Philosophical Quarterly*, 2008), "Contextualism and the Factivity Problem" (*Philosophy and Phenomenological Research*, 2008), and "A Puzzle about Responsibility" (*Erkenntnis*, 2011). Contact: pbauman1@swarthmore.edu.

Dana Maria Bichescu-Burian is PhD from the University of Konstanz in Germany. She is a clinical psychologist specialized in Psychotraumatology and Cognitive-Behavioral Psychotherapy. She is currently a postdoctoral researcher within The Knowledge Based Society Project (POSDRU ID 56815) and a scientific researcher of the University of Ulm and Center for Psychiatry Südwürttemberg, Germany. Her main research area includes studies on traumatic antecedents of psychiatric patients, investigations of consequences of organized violence, and clinical trials. Other topics of her research have been the psychophysiological investigation of traumatized persons and health services studies. Contact: dana.bichescu@gmx.de.

J. Adam Carter is lecturer in philosophy at Queen's University, Belfast. His main research interests are in epistemology, value theory and (more recently) the philosophy of language. He is the author of "Against Swamping" (with Benjamin Jarvis, forthcoming in *Analysis*), "A Problem for Pritchard's Anti-Luck Virtue Epistemology" (*Erkenntnis*, 2011), and "Is Epistemic Expressivism Incompatible with Inquiry" (with Matthew Chrisman, *Philosophical Studies*, 2011). Contact: adam.carter@qub.ac.uk.

Ezio Di Nucci (PhD, Edinburgh) is post-doc in philosophy at Universität Duisburg-Essen. He works mainly in the philosophy of action, free will, normative ethics, and applied ethics. His books include *Mind Out of Action* (2008) and *Content, Consciousness, and Perception* (with Conor McHugh, 2006). His journal articles include: "Simply, false" (*Analysis*, 2009), "Refuting a Frankfurtian Objection to Frankfurt-Type Counterexamples" (*Ethical Theory and Moral Practice*, 2010), "Rational constraints and the Simple View" (*Analysis*, 2010), "Sexual Rights and Disability" (*Journal of Medical Ethics*, 2011), "Frankfurt versus Frankfurt: a new anti-causalist dawn" (*Philosophical Explorations*, 2011), "Automatic Actions: Challenging Causalism" (*Rationality Markets and Morals*, 2011), "Priming Effects and Free Will" (*International Journal of Philosophical Studies*, forthcoming), and "Self-Sacrifice and the Trolley Problem" (*Philosophical Psychology*, forthcoming). Contact: ezio.dinucci@uni-due.de.

David M. Godden (Ph.D. McMaster University, 2004) is an Assistant Professor of Philosophy at Old Dominion University, with research interests in epistemology, the theory of rationality, reasoning and argument, the theory of evidence, the history and philosophy of logic, and 20th century analytic philosophy. He has published on a wide variety of topics including psychologism, Quine's holism, disagreement, common knowledge, presumption and argumentation schemes, and his work has appeared in journals such as *Synthese*, *History and Philosophy of Logic*, *Argumentation*, *Ratio Juris*, *Philosophy & Rhetoric*, *Cogency* and *Informal Logic*. Contact: dgodden@odu.edu.

Susan Haack is Distinguished Professor in the Humanities, Cooper Senior Scholar in Arts and Sciences, Professor of Philosophy, and Professor of Law at the University of Miami. She is the author of many influential books and articles in epistemology, logic and philosophy of science. Her most recent books include *Evidence and Inquiry: A Pragmatist Reconstruction of Epistemology* (2009), *Putting Philosophy to Work: Inquiry and Its Place in Culture* (2008), and *Defending Science-Within Reason: Between Scientism and Cynicism* (2007). Professor Haack is also editor (with Robert Lane) of *Pragmatism, Old and New: Selected Writings* (2006). Contact: shaack@law.miami.edu.

Michael Harbour, Ph.D. in philosophy from Vanderbilt University, is currently enrolled in Harvard Law School. His research focuses on social and political philosophy. He is the author of: "On Epistemic Abstemiousness: a Reply to Bundy" (with Scott F. Aikin, Jonathan Neufeld, and Robert B. Talisse, *Logos & Episteme*, 2011), "Epistemic Martyrs, Epistemic Abstainers, and Epistemic Converts" (with Scott F. Aikin, Jonathan Neufeld, and Robert B. Talisse, *Logos & Episteme*, 2010), "Nagel on Public Education and Intelligent Design" (*Journal of Philosophical Research*, 2010, with Scott F. Aikin and Robert B. Talisse) and "The Ethics of Inquiry and Engagement: The Case of Science in Public" (*Public Affairs Quarterly*, 2010, with Scott F. Aikin). Contact: mharbour@jd13.law.harvard.edu.

Nicholaos Jones is Assistant Professor of Philosophy at University of Alabama in Huntsville. His main areas of interest are philosophy of science and Asian philosophy. His areas of competence are epistemology, logic, and ethics. He is the author of "Diagrams as Locality Aids for Explanation and Model Construction in Cell Biology" (with Olaf Wolkenhauer, *Biology and Philosophy*, 2012), "Nyaya-Vaisesika Inherence, Buddhist Reduction, and Huayan Total Power" (*Journal of Chinese Philosophy*, 2010), and "General Relativity and the Standard Model: Why Evidence for One Does Not Disconfirm the Other" (*Studies in History and Philosophy of Modern Physics*, 2009). Contact: nick.jones@uah.edu.

Jimmy Alfonso Licon is M.A. Candidate in Philosophy, Department of Philosophy, San Francisco State University. His research interests include various aspects of metaphysics (animalist theories of personal identity, time and metaphysical naturalism), epistemology (theories of justification, varieties of scepticism and the problem of easy knowledge) and applied ethics (animal rights/welfare and political legitimacy). He is the author "No Suicide for Presentists: A Reply to Hales" (*Logos & Episteme*, 2011). Contact: jalicon@mail.sfsu.edu.

Franck Lihoreau is a Research Fellow for the Portuguese Science and Technology Foundation at the New University of Lisbon. His main areas of research include epistemology, philosophy of language, philosophical logic, and metaphysics. He has publications in all of these and is the editor of several volumes, including

Knowledge and Questions (Rodopi, 2008) and *Truth in Fiction* (Ontos Verlag, 2011). Contact: franck.lihoreau@fcsh.unl.pt

Jonathan Neufeld is Assistant Professor of Philosophy, College of Charleston. His research interests are in: philosophy of music, aesthetics, political philosophy, and philosophy of law. He is particularly interested in problems surrounding performance and interpretation. He is currently working on two books: *Music in Public: How Performance Shapes Democracy* (forthcoming), *and Listeners, Critics, and Judges* (in preparation). With Scott F. Aikin, Michael Harbour, and Robert B. Talisse, he published in *Logos & Episteme* "On Epistemic Abstemiousness: a Reply to Bundy" (2011) and "Epistemic Martyrs, Epistemic Abstainers, and Epistemic Converts" (2010). Contact: neufeldja@cofc.edu.

Robert B. Talisse is Professor of Philosophy and Political Science and Director of Graduate Studies in Philosophy at Vanderbilt University, and editor of Public Affairs Quarterly. His area of specialization is contemporary political philosophy. His most recent work engages issues at the intersection of political theory and epistemology. He is the author of: *Reasonable Atheism* (with Scott Aikin, 2011), *Democracy and Moral Conflict* (2009), *Pragmatism: A Guide for the Perplexed* (with Scott Aikin, 2008), *A Pragmatist Philosophy of Democracy* (2007), and *Democracy After Liberalism* (2005). In 2010 and 2011, he published in *Logos & Episteme* "Epistemic Martyrs, Epistemic Abstainers, and Epistemic Converts" and "On Epistemic Abstemiousness: a Reply to Bundy" (with Scott F. Aikin, Michael Harbour, and Jonathan Neufeld). Contact: robert.talisse@vanderbilt.edu.

Brian Weatherson is Associate Professor of Philosophy at Rutgers University. His areas of specialization are epistemology and philosophy of language and his areas of competence are metaphysics, philosophy of mind, and logic. He is the author of "Defending Interest-Relative Invariantism (*Logos & Episteme*, 2011), "Deontology and Descartes' Demon" (*Journal of Philosophy*, 2008), "The Bayesian and the Dogmatist" (*Proceedings of the Aristotelian Society*, 2007), and "Can we Do Without Pragmatic Encroachment?" (*Philosophical Perspectives*, 2005). Contact: brian@weatherson.org.

LOGOS & EPISTEME: *AIMS & SCOPE*

Logos & Episteme is a quarterly open-access international journal of epistemology that appears in printed and on-line version at the end of March, June, September, and December. Its fundamental mission is to support and publish various current reflections and researches that aim at investigating, analyzing, interpreting or philosophically explaining the human knowledge in all its aspects, forms, types, dimensions or practices.

For this purpose, the journal will publish articles, reviews or discussion notes focused as well on problems concerning the general theory of knowledge, as on problems specific to the philosophy, methodology and ethics of science, philosophical logic, metaphilosophy, moral epistemology, epistemology of art, epistemology of religion, social or political epistemology, epistemology of communication. Studies in the history of science and of the philosophy of knowledge, or studies in the sociology of knowledge, cognitive psychology, and cognitive science are also welcome.

The journal intends to promote all methods, perspectives and traditions in the philosophical analysis of knowledge, from the normative to the naturalistic and experimental, and from the Anglo-American to the continental or Eastern.

The journal accepts for publication texts in English, French and German, which satisfy the norms of clarity and rigour in exposition and argumentation.

NOTES TO CONTRIBUTORS

1. Accepted Papers

The journal accepts for publication articles, discussion notes and book reviews submitted exclusively to *Logos & Episteme* and not published, in whole or substantial part, elsewhere. The editors of *Logos & Episteme* reserve the right to refuse all future papers belonging to the authors who fail to observe this rule.

2. Submission Address

Please submit your manuscripts electronically at: logosandepisteme@yahoo.com. Authors will receive an e-mail confirming the submission. All subsequent correspondence with the authors will be carried via e-mail. When a paper is co-written, only one author should be identified as the corresponding author.

3. Paper Size

The articles should normally not exceed 12000 words in length, including footnotes and references. Articles exceeding 12000 words will be accepted only occasionally and upon a reasonable justification from their authors. The discussion notes must be no longer than 3000 words and the book reviews must not exceed 4000 words, including footnotes and references. The editors reserve the right to ask the authors to shorten their texts when necessary.

4. Manuscript Format

Manuscripts should be formatted in Rich Text Format file (*rtf) or Microsoft Word document (*docx) and must be double-spaced, including quotes and footnotes, in 12 point Times New Roman font. Where manuscripts contain special symbols, characters and diagrams, the authors are advised to also submit their paper in PDF format. Each page must be numbered and footnotes should be numbered consecutively in the main body of the text and appear at footer of page. For all references authors must use the Humanities style, as it is presented in The Chicago Manual of Style, 15th edition. Large quotations should be set off clearly, by indenting the left margin of the manuscript or by using a smaller font size. Double quotation marks should be used for direct quotations and single quotation marks should be used for quotations within quotations and for words or phrases used in a special sense.

5. Official Languages

The official languages of the journal are: English, French and German. Authors who submit papers not written in their native language are advised to have the article checked for style and grammar by a native speaker. Articles which are not linguistically acceptable may be rejected.

6. Abstract

All submitted articles must have a short abstract not exceeding 200 words in English and 3 to 6 keywords. The abstract must not contain any undefined abbreviations or unspecified references. Authors are asked to compile their manuscripts in the following order: title; abstract; keywords; main text; appendices (as appropriate); references.

7. Author's CV

A short CV including the author`s affiliation and professional address must be sent in a separate file. All special acknowledgements on behalf of the authors must not appear in the submitted text and should be sent in the separate file. When the manuscript is accepted for publication in the journal, the special acknowledgement will be included in a footnote on the first page of the paper.

8. Review Process

The reason for these requests is that all articles, with the exception of articles from the invited contributors, will be subject to a strict blind-review process. Therefore the authors should avoid in their manuscripts any mention to their previous work or use an impersonal or neutral form when referring to it. The review process is intended to take no more than six months. Authors not receiving any answer during the mentioned period are kindly asked to get in contact with the editors. Processing of papers in languages other than English may take longer. The authors will be notified by the editors via e-mail about the acceptance or rejection of their papers. The editors reserve their right to ask the authors to revise their papers and the right to require reformatting of accepted manuscripts if they do not meet the norms of the journal.

9. Acceptance of the Papers

The editorial committee has the final decision on the acceptance of the papers. Papers accepted will be published, as far as possible, in the order in which they are received and they will appear in the journal in the alphabetical order of their authors.

10. Responsibilities

Authors bear full responsibility for the contents of their own contributions. The opinions expressed in the texts published do not necessarily express the views of the editors. It is the responsibility of the author to obtain written permission for quotations from unpublished material, or for all quotations that exceed the limits provided in the copyright regulations. The papers containing racist and sexist opinions assumed by their authors will be rejected. The presence in texts of sexist or racist terms is accepted only if they occur in quotations or as examples.

11. Checking Proofs

Authors should retain a copy of their paper against which to check proofs. The final proofs will be sent to the corresponding author in PDF format. The author must send an answer within 3 days. Only minor corrections are accepted and should be sent in a separate file as an e-mail attachment.

12. Reviews

Authors who wish to have their books reviewed in the journal should send them at the following address: Institutul de Cercetări Economice şi Sociale „Gh. Zane" Academia Română, Filiala Iaşi, Str. Teodor Codrescu, Nr. 2, 700481, Iaşi, România. The authors of the books are asked to give a valid e-mail address where they will be notified concerning the publishing of a review of their book in our journal. The editors do not guarantee that all the books sent will be reviewed in the journal. The books sent for reviews will not be returned.

13. Property & Royalties

Articles accepted for publication will become the property of *Logos & Episteme* and may not be reprinted or translated without the previous notification to the editors. No manuscripts will be returned to their authors. The journal does not pay royalties. Authors of accepted manuscripts will receive free of charge two copies of the issue containing their papers.

14. Permissions

Authors have the right to use their papers in whole and in part for non-commercial purposes. They do not need to ask permission to re-publish their papers but they are kindly asked to inform the Editorial Board of their intention and to provide acknowledgement of the original publication in *Logos & Episteme*, including the title of the article, the journal name, volume, issue number, page number and year of publication. All articles are free for anybody to read and download. They can also be distributed, copied and transmitted on the web, but only for non-commercial purposes, and provided that the journal copyright is acknowledged.

Bun de tipar: 2012 • Apărut: 2012 • Format 17 × 24 cm

INSTITUTUL EUROPEAN

Iaşi, str. Grigore Ghica Vodă nr. 13
Tel. Dif.: 0788.319462; Fax: 0232/230197
editura_ie@yahoo.com; www.euroinst.ro

Volum elaborat în cadrul Proiectului POSDRU/89/1.5/S/56815.
Se distribuie GRATUIT.