Savage, Leonard J. (1954) *The Foundations of Statistics* (New York: Wiley).
——(1971) 'Elicitation of Personal Probabilities and Expectations', *Journal of the American Statistical Association*, 66: 783–801.
——(1972) *The Foundations of Statistics*, 2nd edn. (New York: Dover).
Schervish, Mark J. (1989) 'A General Method for Comparing Probability Assessors', *Annals of Statistics*, 17/4: 1856–79.
Seidenfeld, Teddy (1985) 'Calibration, Coherence, and Scoring Rules', *Philosophy of Science*, 52: 274–94.
Shah, Nishi (2003) 'How Truth Governs Belief', *Philosophical Review*, 112/3 (Oct.): 447–82.
Shuford, E. H., A. Albert, and H. E. Massengill, (1966) 'Admissible Probability Measurement Procedures', *Psychometrika*, 31: 125–45.
Velleman, David (2000) 'On the Aim of Belief', *The Possibility of Practical Reason* (Oxford: Clarendon Press), 244–81.

7. Rationality and Self-Confidence

Frank Arntzenius

1. WHY BE SELF-CONFIDENT?

Hair-Brane theory is the latest craze in elementary particle physics. I think it unlikely that Hair-Brane theory is true. Unfortunately, I will never know whether Hair-Brane theory is true, for Hair-Brane theory makes no empirical predictions, except regarding some esoteric feature of the microscopic conditions just after the Big Bang. Hair-Brane theory, obviously, has no practical use whatsoever. Still, I care about the truth: I want my degree of belief D(H) in Hair-Brane theory H to be as close to the truth as possible. To be precise:

(1) if H is true then having degree of belief $D(H) = r$ has epistemic utility $U(D) = r$ for me
(2) if H is false then having degree of belief $D(H) = r$ has epistemic utility $U(D) = 1 - r$ for me.

Currently, my degree of belief $D(H) = 0.2$. Am I, by my own lights, doing a good epistemic job? Let's see. The expected epistemic utility EU of degree of belief $D'(H) = r$, given my current degree of belief D, is:

(3) $EU(D') = D(H)U(D'\&H) + D(\neg H)U(D'\&\neg H) = 0.2r + 0.8(1 - r) = 0.8 - (0.6)r$.

Obviously, $EU(D')$ is maximal for $D'(H) = 0$. So, by my own lights, I am not doing a good job; I would do better if I were absolutely certain that Hair-Brane theory is false. Unfortunately, I am not capable of setting my degrees of belief at will. All I can do is recognize my own epistemic shortcomings. So I do.

The above is a strange story. Real people, typically, do not judge their own degrees of belief as epistemically deficient. To coin a term: real people tend to be "self-confident." The puzzle that Gibbard poses is that he can see no good reason to be self-confident. For, according to Gibbard, all that follows from having the truth as one's goal, all that follows from having the accuracy of one's state of

belief as one's desire, is that a higher degree of belief in the truth is better than a lower degree of belief in the truth. That is to say, according to Gibbard, the only constraint on the epistemic utilities of a rational person is that they should increase as her degrees of belief get closer to the truth. The simplest, most natural, epistemic utility function ('scoring' function), which satisfies this constraint, is a linear function. In the case of a single proposition, the function that I stated in (1) and (2) is such a function. So, according to Gibbard, not only is it rationally acceptable to judge one's own degrees of belief as epistemically deficient, it is very natural to do so.

In the next section I will suggest that considerations regarding updating can serve to explain why real people are self-confident. However, I will then go on to explain why I am nonetheless sympathetic to Gibbard's suggestion that one cannot give a purely epistemic justification for why our belief states are as they are.

2. UPDATING AND SELF-CONFIDENCE

Gibbard's considerations are entirely synchronic. That is to say, he does not consider the evolution of one's belief state through time. But having the truth as one's goal surely includes the desire to get closer to the truth as time passes. In this section I will try to incorporate such considerations.

Let's start with a simple example. Suppose I initially have the following degree of belief distribution D:

(4) $D(H\&E) = 0.4$
(5) $D(H\&\neg E) = 0.2$
(6) $D(\neg H\&E) = 0.1$
(7) $D(\neg H\&\neg E) = 0.3$

And suppose that I have a linear epistemic utility function. In particular, suppose that, according to my current degrees of belief D, the expected epistemic utility of degree of belief distribution D′ is:

(8) $0.4D'(H\&E) + 0.2D'(H\&\neg E) + 0.1D'(\neg H\&E) + 0.3D'(\neg H\&\neg E)$

This is maximal for $D'(H\&E) = 1$. So, epistemically speaking, I desire that I currently be certain that H&E is true, even though in fact I am not certain of that at all.

Now suppose that I know that in one hour I will learn whether E is true or not. And suppose that the only thing I now care about is the degrees of belief that I will have one hour from now. If that is so, what should I now regard as epistemically the best policy for updating my degrees of belief in the light of the evidence that I will get? That is to say, what degrees of belief D_E do I now think I should adopt if I were to get evidence E, and what degrees of belief $D_{\neg E}$ do I now think I should adopt if I were to get evidence ¬E? Well, my current expected epistemic utility for my future degrees of belief is:

(9) $0.4U(H\&E\&D_E) + 0.2U(H\&\neg E\&D_{\neg E}) + 0.1U(\neg H\&E\&D_E) + 0.3U(\neg H\&\neg E\&D_{\neg E})$.

We can expand each of the four epistemic utilities that occur in (9):

(10) $U(H\&E\&D_E) = D_E(H\&E) + (1 - D_E(H\&\neg E)) + (1 - D_E(\neg H\&E)) + (1 - D_E(\neg H\&\neg E))$
(11) $U(H\&\neg E\&D_{\neg E}) = (1 - D_{\neg E}(H\&E)) + D_{\neg E}(H\&\neg E) + (1 - D_{\neg E}(\neg H\&E)) + (1 - D_{\neg E}(\neg H\&\neg E))$
(12) $U(\neg H\&E\&D_E) = (1 - D_E(H\&E)) + (1 - D_E(H\&\neg E)) + D_E(\neg H\&E)) + (1 - D_E(\neg H\&\neg E))$
(13) $U(\neg H\&\neg E\&D_{\neg E}) = (1 - D_{\neg E}(H\&E)) + (1 - D_{\neg E}(H\&\neg E)) + (1 - D_{\neg E}(\neg H\&E)) + D_{\neg E}(\neg H\&\neg E)$

After substituting these terms into (9) and fiddling around a bit we find that my expected epistemic utility is:

(14) $3 + 0.3D_E(H\&E) - 0.5D_E(H\&\neg E) - 0.3D_E(\neg H\&E) - 0.5D_E(\neg H\&\neg E) - 0.5D_{\neg E}(H\&E) - 0.1D_{\neg E}(H\&\neg E) - 0.5D_{\neg E}(\neg H\&E) + 0.1D_{\neg E}(\neg H\&\neg E)$.

This expression is maximized by setting $D_E(H\&E) = 1$ and $D_{\neg E}(\neg H\&\neg E) = 1$ (and setting the other degrees of belief equal to 0). So if all I care about is the degrees of belief I will have one hour from now, then I should update on E by becoming certain that H&E is true, and I should update on ¬E by becoming certain that ¬H&¬E is true. In particular, by my current lights, it would be wrong to first change my degrees of belief so as to maximize my current expected epistemic utility, then update these degrees of belief by conditionalization, and then change these conditionalized degrees of belief so as to maximize expected epistemic utility by the lights of these conditionalized degrees of belief. So there is a conflict between maximizing the expected epistemic utility of my

current degrees of belief (by my current lights), and maximizing the expected epistemic utility of my future degrees of belief (by my current lights). At least there is such a conflict, if I update by conditionalization.

Given that there is such a purely epistemic conflict, the obvious question is: what should I do if I only have epistemic concerns and I care both about the accuracy of my current degrees of belief and about the accuracy of my future degrees of belief? One might answer: no problem, I should maximize expected epistemic utility (by my current lights) of both of my current degrees of belief and my future degrees of belief, and hence I should jettison conditionalization. That is to say I should now set my degree of belief to $D'(H\&E) = 1$. And then, if I get evidence E, my degrees of belief should stay the same, but if I get evidence $\neg E$, I should set my degrees of belief to $D_{\neg E}(\neg H\&\neg E) = 1$. Unfortunately, there are two problems with this answer.

In the first place, it seems worrying to jettison conditionalization. The worry is not just the general worry that conditionalization is part of the standard Bayesian view. The worry, more specifically, is that if one rejects conditionalization one will have to reject standard arguments in favor of conditionalization, namely diachronic Dutch books arguments. But if one does that, shouldn't one also reject synchronic Dutch book arguments? And if one does that, then why have degrees of belief, which satisfy the axioms of probability, to begin with? I will return to this question in section 4. For now, let me turn to the second problem.

The second problem is that if one were to reset one's current degrees of belief so as to maximize one's current expected epistemic utility, one would thereby lose the ability to set one's future degrees of belief so as to maximize the current expected epistemic utility of those future degrees of belief. Let me explain this in a bit more detail.

According to my current degrees of belief D the epistemically best current degree of belief distribution is:

(15) $D'(H\&E) = 1$
(16) $D'(H\&\neg E) = 0$
(17) $D'(\neg H\&E) = 0$
(18) $D'(\neg H\&\neg E) = 0$

Now, according to my original plan, if I were to learn E then I should update by becoming certain that H&E is true, and if I were to learn

$\neg E$ then I should become certain that $\neg H\&\neg E$ is true. But if I were to replace D by D′ then I would lose the information as to what I should do were I to learn $\neg E$. The reason why I originally desire to update on $\neg E$ by becoming certain that $\neg H\&\neg E$, rather than becoming certain that $H\&\neg E$, is that $D(\neg H/\neg E)$ is higher than $D(H/\neg E)$. But if I were to change D into D′ the relevant information is no longer encoded in my degrees of belief: D′ could have come from a degree of belief D (via expected epistemic utility maximization) according to which $D(\neg H/\neg E)$ is lower than $D(H/\neg E)$, but it could also have come from one according to which $D(\neg H/\neg E)$ is higher than $D(H/\neg E)$. That is to say, if one's epistemic utilities are linear, then maximizing the expected epistemic utility (by one's current lights) of one's degrees of belief can make it impossible to maximize the expected epistemic utility (by one's current lights) of one's degrees of belief at a future time.

The obvious solution to this problem is for the ideal rational agent to have two separate degree of belief distributions. An ideal rational agent should have a 'prudential' degree of belief distribution, which she uses to guide her actions *and* to compute epistemic utilities, and an 'epistemic' degree of belief distribution, which she always sets in order to maximize epistemic utility.

Now, one might worry that there is still going to be a problem. For consider again the example that I started this section with, i.e. suppose that my initial prudential degrees of belief are,

(19) $D^{pr}(H\&E) = 0.4$
(20) $D^{pr}(H\&\neg E) = 0.2$
(21) $D^{pr}(\neg H\&E) = 0.1$
(22) $D^{pr}(\neg H\&\neg E) = 0.3$

Suppose I use these initial prudential degrees of belief to set my initial epistemic degrees of belief so as to maximize expected epistemic utility. Then my initial epistemic degrees of belief would be:

(23) $D^{ep}(H\&E) = 1$
(24) $D^{ep}(H\&\neg E) = 0$
(25) $D^{ep}(\neg H\&E) = 0$
(26) $D^{ep}(\neg H\&\neg E) = 0$

Now I don't (yet) need to worry that I have lost the possibility of maximizing the expected epistemic utility (according to my initial prudential degrees of belief) of my epistemic degrees of belief one

hour from now, since, even though I adopted initial epistemic degrees of belief as indicated, I have retained my initial prudential degrees of belief. However there might still be a problem. For when I acquire evidence E, or evidence ¬E, I will, presumably, update my prudential degrees of belief by conditionalization. So will our problem therefore reappear? Will my *updated* prudential degrees of belief contain enough information for me to be able to deduce from them which epistemic degree of belief distribution has maximal expected epistemic utility according to my *initial* prudential degree of belief distribution? And, even if I do have enough information to be able to stick to my original plan, will that plan still look like a good plan according to my *updated* prudential degrees of belief? Let's see.

Recall that according to my initial prudential degrees of belief, if all I care about is the epistemic utility of my degrees of belief one hour from now, then I should update on E by becoming certain that H&E is true, and I should update on ¬E by becoming certain that ¬H&¬E is true. Now, if I were to learn E and update my prudential degree of belief by conditionalization, then my prudential degrees of belief would become,

(27) $D^{pr}(H\&E) = 0.66$
(28) $D^{pr}(H\&\neg E) = 0.33$
(29) $D^{pr}(\neg H\&E) = 0$
(30) $D^{pr}(\neg H\&\neg E) = 0$

According to these prudential degrees of belief expected epistemic utility is maximized by being certain that H&E is true.

Similarly, if I were to learn ¬E and I conditionalized on this, then my prudential degrees of belief would become,

(31) $D^{pr}(H\&E) = 0$
(32) $D^{pr}(H\&\neg E) = 0$
(33) $D^{pr}(\neg H\&E) = 0.75$
(34) $D^{pr}(\neg H\&\neg E) = 0.25$

According to these prudential degrees of belief expected epistemic utility is maximized by being certain that ¬H&¬E is true.

So, in this case at least, the epistemic degrees of belief that I should adopt in the light of evidence, according to my initial prudential degrees of belief, are the same as the ones that I should adopt according to my later prudential degrees of belief, if I update my prudential degrees of belief by conditionalization.

What it is more interesting, and perhaps more surprising, is that this is true for every possible initial prudential degree of belief distribution, *and for every possible epistemic utility function*. That is to say, no matter what one's epistemic utilities are, if according to one's prudential degrees of belief at some time t, plan P for updating one's epistemic degrees of belief maximizes expected epistemic utility, then, after one has updated one's prudential degrees of belief by conditionalization, plan P will still maximize expected utility according to one's updated prudential degrees of belief. The proof of this fact for the general finite case is simple, so let me give it.

Let $D^{pr}(W_i)$ be my initial prudential degree of belief distribution over possibilities W_i.[1] Let $U(W_i\&D^{ep})$ be my epistemic utility for having degree of belief distribution D^{ep} in possibility W_i. Suppose I know that in an hour I will learn which of $E_1, E_2, \ldots E_n$ is true (where the E_i are mutually exclusive and jointly exhaustive). An 'epistemic plan' P is a map from current prudential degree of belief distributions plus evidence sequences to future epistemic degree of belief distributions. Let P map D^{pr} plus E_i to $D^{ep}{}_i$. Then P has maximal expected epistemic utility according to D^{pr} and U iff for every alternative plan P′ (which maps D^{pr} plus E_i to $D^{ep\prime}{}_i$) we have:

(35) $\sum_i \sum_k D^{pr}(W_k\&E_i)U(W_k\&E_i\&D^{ep}{}_i) \geq \sum_i \sum_k D^{pr}(W_k\&E_i)U(W_k\&E_i\&D^{ep\prime}{}_i)$

We can rewrite this as,

(36) $\sum_i \sum_k D^{pr}(E_i)D^{pr}(W_k/E_i)U(W_k\&E_i\&D^{ep}{}_i) \geq \sum_i \sum_k D^{pr}(E_i)D^{pr}(W_k/E_i)U(W_k\&E_i\&D^{ep\prime}{}_i)$.

The left-hand side being maximal implies that each separate i-term is maximal:

(37) $\sum_k D^{pr}(E_i)D^{pr}(W_k/E_i)U(W_k\&E_i\&D^{ep}{}_i) \geq \sum_k D^{pr}(E_i)D^{pr}(W_k/E_i)U(W_k\&E_i\&D^{ep\prime}{}_i)$, for each i.

Therefore,

(38) $\sum_k D^{pr}(W_k/E_i)U(W_k\&E_i\&D^{ep}{}_i) \geq \sum_k D^{pr}(W_k/E_i)U(W_k\&E_i\&D^{ep\prime}{}_i)$, for each i.

[1] I am assuming that my degrees of belief are not part of the possibilities W_i that I distribute my degrees of belief over.

But this just means that if we conditionalize D^{pr} on E_i then, according to the resulting degree of belief distribution (and U), the expected epistemic utility of $D^{ep}{}_i$ is maximal.

Let me now summarize what we have seen in this section, and draw a tentative conclusion. No matter what one's epistemic utility function is, one can maximize one's epistemic utilities at all times by having two separate degree of belief distributions: a prudential degree of belief distribution which guides one's actions and one's choice of an epistemic degree of belief distribution, and an epistemic degree of belief distribution whose sole purpose is to maximize epistemic utility. One can then give a purely epistemic argument for updating one's prudential degrees of belief by conditionalization, on the grounds that such updating guarantees cross-time consistency of epistemic utility maximization. The epistemic degrees of belief of an ideal agent at a given time do not determine how she updates her epistemic degrees of belief in the light of evidence. Rather, she updates her epistemic degrees of belief by first conditionalizing her prudential degrees of belief and then maximizing epistemic utility. Thus the epistemic degrees of belief of an ideal agent are largely epiphenomenal: they are only there to maximize the *epistemic score* of an agent, they are not there to guide her actions, nor are they there to help determine her future epistemic degrees of belief. This suggests that rational people can make do without epistemic utilities and epistemic degrees of belief, which could explain why real people do not consider themselves epistemically deficient. Let me bolster this suggestion by arguing that it is not clear what epistemic utilities are.

3. WHAT ARE EPISTEMIC UTILITIES?

Gibbard characterizes epistemic utilities, roughly, as follows. Person P's epistemic utilities are the utilities that P would have were P to ignore both the 'guidance' value and the 'side' values of his degrees of belief. The 'guidance' value of P's degrees of belief is the value these degrees of belief have for P due to the way in which they guide P's actions. The 'side' values of P's degrees of belief for P are values such as P's happiness due to, for example, P's certitude that he will have a pleasant afterlife, or P's dejection due to, for example, P's certitude of his own moral inferiority, and so on. My worry

now is that it is not clear what epistemic utilities are, and hence it is not clear that rational people must have epistemic utilities. That is to say, I am willing to grant that rational people have all-things-considered utilities. But it is not clear to me exactly what should be 'subtracted' from 'all considerations' in order to arrive at purely 'epistemic' utilities.

Consider, for instance, my home robot servant, Hal. The robot factory equipped Hal with reprogrammable degrees of belief, reprogrammable utilities, a conditionalization module, and an expected utility maximization module. When I bought Hal I set his degrees of belief equal to mine, his utilities equal to mine (that is to say, my 'all-things-considered' utilities), and I instructed Hal to act on my behalf when I was not present. Occasionally Hal and I updated each other on the evidence that each of us received since our last update, and all went well. Unfortunately Hal's mechanics broke down a while ago. That is to say Hal still has degrees of belief and utilities, and can still conditionalize and compute expected utilities, but he can no longer perform any actions. He just stands there in the corner, a bit forlorn. I have not bothered updating Hal recently, since he can't do anything any more. Gibbard asks me, "Suppose you just wanted Hal's current degrees of belief to be accurate, what degrees of belief would you give him?" I answer, "I don't know. Tell me what you mean by the word 'accurate', and I will tell you what I would set them to." For instance, suppose that there is only one proposition p that Hal has degrees of belief in. Of course if I know that p is true, then I will judge Hal's degrees of belief the more accurate the higher Hal's degree of belief in p is. That much presumably follows from the meaning of the word "accurate." But this by itself does not determine what I take to be the accuracy of Hal's degrees of belief when I am uncertain as to whether p is true or not. Nor does it even allow me to figure out the *expected* accuracy of Hal's degrees of belief. In order to be able to calculate such *expected* accuracies, I need to attach *numerical* values to the accuracy of degree of belief distribution/world pairs (where these numerical values are unique up to positive linear transformations). And I don't know how to do that. So I am stuck. I suggest that this is not for lack of rationality or lack of self-knowledge on my part, but rather, because Gibbard is asking an unclear question. Presumably Gibbard would respond that the above paragraph is

confused. On his view of course the question, "What would you set Hal's degrees of belief to if you just wanted them to be accurate?" does not have a person-independent, objectively correct, answer. The problem, according to Gibbard, is precisely that one could rationally have epistemic utilities such that one desires to set Hal's degrees of belief to equal one's own degrees of belief, but one's epistemic utilities could also be such that one desires to set Hal's degrees of belief to be different from one's own degrees of belief. This just goes to show that the correct answer to his question is person-dependent.

My worry, however, is not that Gibbard's question is a well-defined question which has a person-dependent answer, but rather that his question is not a well-defined question. My worry is that it is a question like, "What color socks do you want Hal to wear, bearing in mind that your only goal is colorfulness?" I can't answer that question, not because I am not clear about my own desires or because I am not rational, but because the term "colorfulness" is too vague, or ill-defined. Similarly, I worry that the term "epistemic" is too vague, or ill-defined, so that there are no well-defined (person-dependent) numerical epistemic utilities.

4. WHY HAVE DEGREES OF BELIEF?

Suppose one's only concerns are epistemic. Why then have degrees of belief? That is to say, when one's only goal is truth why should one's epistemic state satisfy the axioms of probability theory? I see no good reason. Let me indicate why I am skeptical by very briefly discussing standard arguments for having belief states which satisfy the axioms of probability theory.

Standard Dutch book arguments rely on the assumption that one does not want to be guaranteed to lose money, or, more generally, that one does not want to be guaranteed to lose prudential value. So, prima facie, if one's only concerns are epistemic, Dutch book arguments have no bite. However, there have been attempts to remove prudential considerations from Dutch book arguments. (See, for instance, Howson and Urbach 1989; Hellman 1997; or Christensen 1996.) The basic idea of these attempts is to claim that the epistemic states of rational people must include judgments regarding the 'fairness' of bets, where these judgments have to satisfy certain axioms

which, in turn, entail the axioms of probability theory, so that, purportedly, the epistemic states of rational people must include degrees of belief which satisfy the axioms of probability theory.

There are two reasons why such arguments do not show that one's epistemic state must include degrees of belief which satisfy the axioms of probability theory when one's only goal is the pursuit of truth. In the first place the authors give no justification based only on the pursuit of truth for why epistemic states should include judgments of the 'fairness' of bets. (This may not be a slight on the cited authors, since it is not clear that they intended to give such a justification.) Secondly (and this is a slight on the authors), as argued in Maher (1997), even if a rational person does have epistemic reasons for having such a notion of "fairness" of bets, the authors' arguments for why this notion should satisfy the suggested axioms are not convincing. In fact, Maher shows that some of the suggested axioms will typically be violated by rational people. For instance, if a person judges a bet to be fair just in case the expected utility of accepting the bet is zero, and if her utilities are non-linear in dollars, then her judgments of fairness will violate some of the proffered axioms.

The next type of arguments rely on so-called 'representation theorems.' Such theorems show that preferences which satisfy certain axioms are always representable as those of an expected utility maximizer who has degrees of belief which satisfy the axioms of probability theory. I already find it hard to see why a rational person's all-things-considered preferences should satisfy some of these axioms.[2] I find it even harder to see why a person's purely epistemic preferences should do so, even assuming that sense can be made of 'purely epistemic' preferences. Let me explain in slightly more detail why I find it so hard to see why there should be purely epistemic preferences which satisfy the axioms needed for representation theorems.

One of the axioms needed for representation theorems is that preferences are transitive: if a rational person prefers A to B and B to C then she prefers A to C. When it comes to all-things-considered preferences this axiom seems to me very plausible. For, on a very

[2] For instance, Jeffrey's *continuity axiom* and Savage's *P6 axiom* seem to have no obvious justification other than mathematical expediency. See Jeffrey (1983: ch. 9), and Savage (1972: ch. 3).

plausible understanding of what all-things-considered preferences are, one can be money pumped if one violates this axiom. Now, however, let us consider the case of purely epistemic preferences. Perhaps in this case too one can be money pumped. Fine, but why should one care if one only has epistemic concerns? One might respond that the money pumping argument should not, at bottom, be taken to be a pragmatic argument which only applies to people who are concerned at avoiding a guaranteed loss of money; rather, the argument serves to demonstrate the fundamental incoherence of preferences which are not transitive. I am not moved by such a reply. It may well be that preferences cannot coherently be taken to violate transitivity. However, that merely shifts the issue. For then the question becomes, "is there any reason for a rational person with purely epistemic concerns to have preferences at all?" I can see no such reason.

Finally, there are arguments such as Cox's theorem, and de Finetti's theorem, which show that "plausibility" judgments which satisfy certain axioms are uniquely representable as numerical degrees of belief which satisfy the axioms of probability theory.[3] Again, I can think of no non-question-begging reason why the epistemic states of rational people with purely epistemic concerns should include "plausibility" judgments which satisfy the axioms in question. Let me give a little bit more detail.

De Finetti's theorem and Cox's theorem do roughly the following: they show that one can recover the quantitative values of a probability distribution from the associated comparative qualitative probability judgments. Now, there is a way in which these theorems are not that surprising. For instance, imagine a probability distribution as represented by a heap of mud lying over a continuous space. Then one can think of the qualitative probability judgments as being claims of the form, "the amount of mud over area A is bigger or smaller than the amount of mud over area B." Now, clearly, one cannot shift the mud around in any way without altering some such qualitative judgments. So the qualitative judgments determine the quantitative probabilities. While this argument as it stands is not precise, and does not prove exactly what de Finetti and Cox proved, it does give one some of the flavor of their theorems.

[3] See, for instance, Jaynes (2003: ch. 2), or Howson and Urbach (1989: ch. 3). The fundamental notions in the case of Cox are "plausibilities" and "conditional plausibilities," and in the case of De Finetti the fundamental notion is that of "comparative likelihood."

Now, while the axioms in question may seem plausible to many, this, it seems to me, is due to the fact that one has in mind that the plausibility assessments are the natural qualitative judgments associated with quantitative probabilistic assessments. One way or another, for instance, the presupposition is made that the possible epistemic states with respect to a single proposition form a one-dimensional continuum, and no argument for this is given based on purely epistemic concerns. More generally, in so far as one thinks that the axioms on plausibility judgments cannot coherently be violated by a rational person with only epistemic concerns, I can see no reason why the epistemic state of a rational person with only epistemic concerns should include such judgments. So Cox's theorem and De Finetti's theorem do not seem to supply a purely epistemic justification for having degrees of belief satisfying the axioms of probability theory.

In short, I am not aware of any good *purely epistemic* argument for having belief states which satisfy the axioms of probability theory. Now, one might respond that, indeed, the reason for having belief states that satisfy the axioms of probability theory is (at least partly) prudential, but that, given that one has such belief states, one can ask whether rational people can have purely epistemic reasons to be dissatisfied with the degrees of belief that they have. However, if a rational person has no purely epistemic reason to have degrees of belief, why think a rational person must have purely epistemic preferences over all possible degree of belief distributions?

5. CONCLUSIONS

The notion of purely epistemic concerns is unclear to me. In so far as it is clear to me I find it hard to see a purely epistemic reason for a rational person to have belief states which satisfy the axioms of probability. If I nonetheless grant that a rational person does have such belief states and that it is clear what purely epistemic concerns are, then I can see reasons for a rational agent to have two different sets of degrees of belief: epistemic ones which serve only to maximize her epistemic utilities, and prudential ones to do everything else. Prudential degrees of belief should then be

updated by conditionalization. Epistemic degrees of belief will get dragged along by the prudential ones, relegating epistemic utilities and epistemic degrees of belief to the status of an unimportant side-show.

REFERENCES

Christensen, David (1996) 'Dutch-Book Arguments Depragmatized: Epistemic Consistency for Partial Believers', *Journal of Philosophy*, 93: 450–79.

Hellman, Geoffrey (1997) 'Bayes and Beyond', *Philosophy of Science*, 64: 191–221.

Howson, Colin, and Peter Urbach (1989) *Scientific Reasoning, the Bayesian Approach* (La Salle, Ill.).

Jaynes, Edwin Thompson (2003) *Probability Theory, the Logic of Science* (Cambridge).

Jeffrey, Richard C. (1983) *The Logic of Decision* (Chicago).

Maher, Patrick (1997) 'Depragmatized Dutch Book Arguments', *Philosophy of Science*, 64: 291–305.

Savage, Leonard J. (1972) *The Foundations of Statistics* (New York).

8. A Note on Gibbard's "Rational Credence and the Value of Truth"

Eric Swanson

Gibbard observes that "With an epistemically rational person, it is as if, by her own lights, she were aiming at truth," and argues that although aiming at guidance value is sufficient for this kind of "epistemic immodesty," aiming at truth alone is not. His arguments trade on an analogy between a certain kind of idealized believer and ordinary believers like us: if it is (in certain respects) "as if" we are such idealized believers, then there is good reason to think that we have (certain of) their features. Here I try to undermine Gibbard's case by showing that for another kind of idealized believer—a kind that is more like us than Gibbard's believers are—in many cases having the aim of truth alone does suffice for epistemic immodesty. In particular, a believer who 'aims at truth' in part by being sensitive to new evidence in the way that is most conducive to the *eventual* accuracy of her beliefs most prefers her actual credences. I don't think this conclusively shows that our having the aim of truth suffices for epistemic immodesty. But it does make me suspect that Gibbard's conclusion that guidance value plays a special role in securing our epistemic immodesty is an artefact of his choice of idealization.

1

Let $g_1(\cdot)$ be a function from a believer's credence in some proposition S to her value for having that credence if S is true, and let $g_0(\cdot)$ be a function from the believer's credence in S to her value for having that credence if S is false. Gibbard says that a believer's valuing

Thanks to Frank Arntzenius, Allan Gibbard, Alan Hájek, and Sarah Moss for discussion and helpful comments on earlier drafts.

"truth and truth alone in her credence in S ... seems to consist in satisfying"

CONDITION $\mathcal{T}$: $g_1(x)$ strictly increases with x increasing, and $g_0(x)$ strictly increases with x decreasing.

In many respects Condition $\mathcal{T}$ is not a substantive constraint. Note, for example, that for any positive m and n it is satisfied by the value functions

$$g_1(x) = x^m$$
$$g_0(x) = 1 - x^n$$

The claim that valuing truth alone is compatible with *such* a wide range of pairs of value functions should be controversial. But this is not to say that Condition $\mathcal{T}$ is toothless. Indeed, I think some argument is needed to show that the value functions of a believer who values truth and truth alone must be *strictly* monotonic, as Condition $\mathcal{T}$ demands. Consider for example a believer who, as her known last act, chooses credences that will maximize expected epistemic value by the lights of the value functions

$$\hat{g}_1(x) = \begin{cases} 1 \text{ if } x = 1; \\ 0 \text{ otherwise} \end{cases}$$

$$\hat{g}_0(x) = \begin{cases} 1 \text{ if } x = 0; \\ 0 \text{ otherwise} \end{cases}$$

Has such a believer ipso facto ceased to value truth? To be sure, she values correct *guesses* at S's truth value while disvaluing accurate estimates of its truth-value, in the sense of Jeffrey (1986) and Joyce (1998). But I find it plausible enough that choosing known-to-be-final credences that are not 'lukewarm' can count as a way of aiming at truth alone.

At any rate, Gibbard thinks that epistemic rationality puts far more substantive constraints on credal value functions. In particular, he thinks that for an epistemically rational agent $g_1(\cdot)$ and $g_0(\cdot)$ must be a **credence eliciting pair**, where this means that a believer with such value functions most prefers to have the credence she actually has. I will say that such a believer is **epistemically immodest** with respect to S. Many pairs of value functions satisfy Condition $\mathcal{T}$ without being credence eliciting, and indeed many plausible strengthenings of Condition $\mathcal{T}$ admit non-credence eliciting pairs of value functions.

For example, one might think that for a believer who values truth alone in her credences must have symmetric value functions, in the sense that for any $x \in [0,1]$, $g_1(x) = g_0(1-x)$.[1] After all, the value of believing S if S is true *just is* the value of disbelieving $\overline{S}$ if $\overline{S}$ is false, and it seems plausible that ways of valuing pure credal accuracy should not be sensitive to the particular proposition that is believed or disbelieved. To motivate this idea in a slightly different way, perhaps "Belief aims at truth" is a special case of the less homey truism that credence aims at accuracy. And a valuation of credal accuracy should not arbitrarily privilege credence in truths or credence in falsehoods by valuing them asymmetrically.

We would then have

CONDITION $\mathcal{T}$, SECOND PASS:

- $g_1(x)$ strictly increases with x increasing, and $g_0(x)$ strictly increases with x decreasing;
- for all $x \in [0,1]$, $g_1(x) = g_0(1-x)$.[2]

One non-credence eliciting pair of value functions that satisfies Second Pass is

$$g_1(x) = x$$
$$g_0(x) = 1 - x$$

As Gibbard notes, a believer with this pair of credal value functions would maximize her expected value by "making [her] beliefs extreme in their certitude" unless her initial credence in S is 0.5. So even Second Pass is satisfied by pairs of value functions that are not credence eliciting.

[1] See Winkler (1994) for some discussion of this sort of symmetry.

[2] I also assume henceforth that credal value functions are well-defined and continuous over $[0,1]$. I think this assumption does need some argument, strictly speaking, but I doubt Gibbard would contest it.

2

Pairs of credal value functions that make the counterintuitive prescription that we set any credence besides 0.5 to one of the extreme values of 0 and 1 are in some intuitive sense credally pernicious. A believer with such values will, if she can, at a given time choose credences that dramatically misrepresent the evidence that she has in fact acquired to that time.

There are several factors that together constitute the credal perniciousness of these particular value functions, however, and it is important to pull them apart. The **report relation** for a pair of credal value functions $g_1(\cdot)$ and $g_0(\cdot)$ is that relation R such that $\alpha R x$ iff, according to $g_1(\cdot)$ and $g_0(\cdot)$, given initial credence α in S, having credence x in S (or reporting credence x in S) maximizes expected value.[3] The report relation for $g_1(x) = x$ and $g_0(x) = (1 - x)$ is:

$$\alpha R \begin{cases} 1 \text{ if } \alpha > 0.5; \\ 0.5 \text{ if } \alpha = 0.5; \\ 0 \text{ if } \alpha < 0.5 \end{cases}$$

R has three properties that encapsulate the credal perniciousness of $g_1(\cdot)$ and $g_0(\cdot)$:

1. Some $\alpha \neq \beta \in [0,1]$ bear R to the same x. R thus *conflates* prior credences.
2. Some values in (0,1) bear R to 0, and some bear R to 1. So applying R to a regular credence distribution will sometimes result in an irregular distribution. Even if regularity in one's credences is not a necessary condition for rationality, it is counterintuitive to value irregularity on purely epistemic grounds.
3. Some values in [0,1] do not bear R to themselves. This is just what it means to have a pair of credal value functions that is not credence eliciting.

[3] It is important to think in terms of report relations instead of report functions because for some credal value functions distinct credences in S yield maximal expected value given a single initial credence. For example, for $g_1(x) = x^2$ and $g_0(x) = (1 - x)^2$, if $\alpha = 0.5$ we have maximal expected value at both $x = 1$ and $x = 0$.

The first two properties mentioned above are artefacts of the particular non-credence eliciting value functions we are considering. So it will be helpful to consider pairs that satisfy Second Pass and exhibit only the third property.

Consider for example the following pair of credal value functions, superficially similar to those for the Brier score.

$$g_1(x) = 1 - (1 - x)^3$$
$$g_0(x) = 1 - x^3$$

This pair has the report relation plotted in Figure 8.1:

$$\alpha R x \text{ iff } \alpha = \frac{x^2}{2x^2 - 2x + 1}$$

For all $\alpha, \beta \in [0,1]$, $\alpha = \beta$ iff α and β bear R to the same x. Moreover, no values between 0 and 1 bear R to 0 or 1. Nevertheless, this pair of credal value functions is not credence eliciting: for every value of α but 0, 0.5, and 1, R is not reflexive.

Can a believer aim at truth, choose credences partly on the basis of these value functions, and update those very credences as new evidence comes in? If a believer is certain that she will get no more evidence that will interact with her level of credence in S—as it were, if she knows that she is on her deathbed and

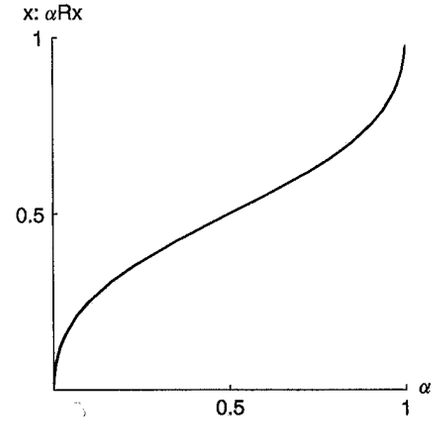

Figure 8.1. The report relation for $g_1(x) = 1 - (1 - x)^3$, $g_0(x) = 1 - x^3$

for whatever reason wants to hedge her bets, somewhat, with her final credences—I think she could choose credences on the basis of this pair of value functions and still count as aiming at the truth. But the circumstances in which a believer can count as aiming at the truth and shift her credences in this way—without making compensating shifts in her updating procedures—are quite rare. A believer who aims at truth alone in her beliefs and thinks that she might get evidence that will interact with her level of credence in S should take every care *not* to let new evidence directly interact with credences that misrepresent her old evidence. Otherwise she would be distorting her total evidence in a way that would undermine her aim to *eventually* estimate the truth in a way that makes the best use of the evidence available to her.

To see this consider a meteorologist trying to decide what value to report as the probability that a storm will pass over a particular island. She is confident that she updates well on the basis of new information, but for prudential reasons she believes that the greater the probability that the storm will pass over the island, the more she should exaggerate that probability in her report. Imagine that she can handle new information that she acquires using either of the following step-by-step strategies:

Applying R at the end

1. She begins to construct an array, writing her initial credences in the first row.
2. When new information comes in, she writes, in row $n+1$ under the last complete row n, the values that would be the product of her updating on that information if her priors were given by row n.
3. For her forecast she reports the relevant value of the image of the last complete row under a non-credence eliciting report relation R.[4]

[4] In these examples suppose that any $a \in [0,1]$ bears R to exactly one x. That is, suppose that the report relation can be understood as a report function that is well-defined over [0, 1].

Applying R with each update

1. She begins to construct an array, writing her initial credences in the first row.
2. She writes, in row $n+1$ under the last complete row n, the image of row n under R.
3. When new information comes in, she writes, in row $n+1$ under the last complete row n, the values that would be the product of her updating on that information if her priors were given by row n, *and then returns to step 2.*
4. For her forecast she reports the relevant value of the last complete row.

Clearly our meteorologist should adopt the first strategy: she should apply R only at the end, so that it affects her report exactly once. Her aim is to estimate the probability that the storm will pass over the island as accurately as she can and then to determine what probability she should report in order to maximize expected value by the lights of her value functions. The second strategy has the potential to lead her far astray—for example, it would make her reported value sensitive to *how many times* she had updated, and this is no part of the way she values estimates of truth value. In brief, the fact that she values estimates of the truth value of a proposition disproportionately depending on their proximity to 1 does not entail that she similarly values what amounts to disproportionate *updating*.

For just these reasons, a believer with non-credence eliciting value functions who aims to eventually estimate the truth in a way that makes the best use of her evidence must have doxastic policies that allow her at least to *emulate* the first strategy. But there is a problem here: a believer with such value functions cannot simply *wait* to apply R until the moment of report, as the meteorologist could. A believer with non-credence eliciting value functions applies R to her credences whenever she acts so to maximize expected value. What constraints does this put on her credal value functions?

Perhaps it would be enough that their report relation be injective. Then one R-step 'backward' from the credences the believer arrived at after taking one R-step 'forward' from her initial credences would return her to her initial credences. The believer's updating procedures could then be modified to update not her actual credences,

but their image under the inverse relation of R. She could in effect 'undo' her previous choice with each update. We would then have the following constraint on the value functions of a believer who aims at truth, thinks she may acquire more evidence, and wants to put that evidence to fruitful epistemic use:

CONDITION τ, THIRD PASS:

- $g_1(x)$ strictly increases with x increasing, and $g_0(x)$ strictly decreases with x decreasing;
- for all $x \in [0,1]$, $g_1(x) = g_0(1-x)$;
- the report relation for $g_1(\cdot), g_0(\cdot)$ is injective over $[0, 1]$.

But this is just more grist for Gibbard's mill, because even this very strong condition does not rule out non-credence eliciting report relations. For example, the report relation of the 'Brier cubed' score is injective over [0, 1], and the relevant pair of value functions satisfies the other clauses of Third Pass.

3

There is a problem with this proposal, however, that points the way toward a constraint that permits only credence eliciting value functions. The doxastic lives of the believers that we have chosen to theorize about consist of sequences of updating and acting so as to maximize expected value. Consider some such believer, who has a non-credence eliciting value function and at some time or times

1. Acts, *inter alia* choosing new credences;
2. Immediately acts again, *inter alia* again choosing new credences, partly on the basis of the credences she chose in the immediately preceding action.

For such a believer to 'work backward' to credences that aren't distorted by non-credence eliciting choices, modifying updating procedures to compensate for non-credence eliciting choices of credences, is not enough: she also needs to ensure that her *actions*—and in particular, her choices of credence—compensate for her choices of credence. Otherwise the believer will choose a credence for a proposition on the basis of a credence which itself may have been chosen to maximize expected value, which itself may have been chosen to maximize expected value, and so on. In virtue

of choosing credences in this fashion—one choice immediately following another—such a believer embodies a one-dimensional discrete dynamical system, the evolution rule of which is $R(\cdot)$. A believer who engages in just one choice of credence takes one step in the dynamical system, so that if she starts with a credence in S of $\rho(S)$, she chooses a credence of $R(\rho(s))$ in S. A believer who engages in two immediately consecutive choices of credence takes two steps, so that she ultimately chooses a credence of $R(R(\rho(s)))$, and so on.

Being able to 'work backward' from the output of such a dynamical system requires much more than that the report relation be injective. For example, it is necessary (though obviously not sufficient) that at least one of the following conditions obtains:

1. The believer knows whether the evolution rule has been applied once or twice.
2. Whether the evolution rule has been applied once or twice doesn't make a difference to how the believer should work backward.

The cognitive lives of believers who satisfy the first condition must be quite transparent to them: they must be able to determine, through introspection, how many times they have acted to maximize expected value. We are unlike such believers in a host of ways. We are *less* unlike believers who do not enjoy such introspective transparency, and thus satisfy only the second condition. To be sure, we are unlike them in important respects as well. But we are *closer* to them than we are to believers who can survey their expected value maximizing actions in the ways necessary to satisfy the first condition.

For a believer to satisfy the second condition, her report relation must not conflate distinct credences under iteration. By this I mean that there must not be distinct credences such that one is related to a value by *one* R-step that the other is related to by *two* R-steps. The only report relation with this feature maps each element in $[0,1]$ to itself and only to itself. And only credence eliciting pairs of credal value functions have this report relation. More formally: The credal value functions $g_1(\cdot)$ and $g_0(\cdot)$ of a believer who thinks she may get new, relevant evidence, and aims at the *eventual* truth—and thus aims to be optimally sensitive to new evidence as it comes in—must satisfy

CONDITION τ, FOURTH PASS: The report relation for $g_1(\cdot)$ and $g_0(\cdot)$ is such that for no $x \neq y \in [0,1]$ is there any z such that xR^2z and yR^1z.[5]

A pair of credal value functions is credence eliciting iff the pair satisfies Fourth Pass.

$\Rightarrow$ Suppose a pair is credence eliciting. Then xRx for any $x \in [0,1]$, and if xRy then $y = x$. So for any k, xR^kx, and if xR^ky then $y = x$. So for any $x \neq y$ and any k and l, x does not bear the R^k relation to any value that y bears the R^l relation to. In particular, x does not bear the R^2 relation to any value that y bears the R^1 relation to.

$\Leftarrow$ Suppose a pair of credal value functions, $g_1(\cdot)$ and $g_0(\cdot)$, satisfies Fourth Pass. Then for no $x \neq y \in [0,1]$ is there any z such that xR^2z and yR^1z. [0, 1] is compact, and $g_1(\cdot)$ and $g_0(\cdot)$ are continuous over [0, 1], so by the extreme value theorem every value in [0, 1] bears the R relation to some value in [0, 1]. In particular, every value in [0, 1] bears the R relation only to itself. For suppose not: then for some $x \neq y \in [0,1]$, xR^1y. There is also some $z \in [0,1]$ such that yR^1z. But then xR^2z, contradicting our initial supposition. So xRx for any $x \in [0,1]$, and if xRy then $y = x$. So the pair is credence eliciting.

This shows that we are more like believers who (in order to aim at eventual truth) must have credence eliciting value functions than we are like believers who can aim at eventual truth without having credence eliciting value functions.

4

How should what we learn about hypothetical believers (who always act to maximize expected value, can choose their own credences, and so on) inform our thinking about believers like us? One reason to think about hypothetical believers in general is that they—or at any rate, some of them—help provide tractable and

[5] aR^nb iff b is accessible from a by an n length sequence of R-steps. So aR^1b iff aRb; aR^2b iff there is some c such that aRc and cRb, etc.

not too misleading models of believers like us. In light of this it would be interesting to see examples of believers whose cognitive lives *demand* that they be modeled using non-credence eliciting value functions: believers for whom it really is "as if" they choose their own credences according to such functions. I am not sure that there are any such believers because, as I have tried to bring out, a believer who has non-credence eliciting value functions and can choose his own credences engages in very odd doxastic behavior over time.

But even without such examples, clarifying the constraints that govern the spaces of various kinds of purely hypothetical believers can point the way toward interesting hypotheses about non-idealized believers. Studying believers that are unlike us can be misleading, however, if we do not correct for artefacts generated by the particular kind of hypothetical believer we choose to focus on. I have argued that for believers who cannot survey the number of times they have acted to maximize expected value, having the aim of eventual truth suffices to ensure that their value functions are credence eliciting. Of course this is compatible with the claim that for *another* kind of believer the aim of eventual truth does not so suffice. But because the believers Gibbard focuses on are in important respects more unlike us than the kind I have discussed, I am not moved to think that it is the aim of maximizing prospective guidance value that secures our epistemic immodesty.

REFERENCES

Jeffrey, Richard C. (1986) 'Probabilism and Induction', *Topoi*, 5: 51–8.

Joyce, James M. (1998) 'A Nonpragmatic Vindication of Probabilism', *Philosophy of Science*, 65(4): 575–603.

Winkler, Robert L. (1994) 'Evaluating Probabilities: Asymmetric Scoring Rules', *Management Science*, 40(11): 1395–405.

9. Aiming at Truth Over Time: Reply to Arntzenius and Swanson

Allan Gibbard

I want to thank both Frank Arntzenius and Eric Swanson for their fine, illuminating commentaries. Both propose that my analysis of belief should be made dynamic. In my paper I considered only a simplest possible case, the static case with a single uncertain proposition and its negation. I might have gone on to consider a more complex thinker, prone to change degrees of credence as new evidence comes in. I agree with Arntzenius and Swanson that the dynamic case needs investigating. I think, however, that for the dynamic case, most of the lessons I drew reappear in new forms.

1. THE PROBLEM

In my paper, recall, I tried to make sense of the idea that "belief aims at truth." I considered epistemic rationality, and asked whether it can somehow be explained as answering to a pure concern with truth. By epistemic rationality, I mean rationality in one's degrees of credence. (For short, following David Lewis, I call degrees of belief "credences.") The epistemic rationality of a state of belief is different from its overall desirability. It is not the same thing as rationality in acting to affect the belief state, or the belief state's being the kind one might go for given the choice. The upshot of my inquiry was both negative and positive. Concern for truth, I first argued, might take any of various forms. Some of these are friendly to epistemic rationality, and some are not. In arguing this, I took for granted standard ideas of what epistemic rationality consists in—or at least, I took the standard decision-theoretic conditions as necessary for perfect epistemic rationality. Concern for truth as such, I thought I showed, couldn't explain epistemic rationality. Epistemic rationality answers to a concern for truth only if the concern takes a special form: that of concern with truth for the sake of guidance.[1]

I helped myself to standard requirements on credence in order to see if they are self-endorsing. Since I took these requirements as assumptions, no argument of the kind I gave could possibly convince anyone of these requirements who wasn't already convinced. Jim Joyce undertakes a more ambitious kind of argument, one that addresses a person who doubts that epistemic rationality requires standard coherence in one's credences—where "standard coherence," as I'm using the term, amounts to satisfying the usual axioms of probability. Joyce tries to show, on the basis of things that such a person would accept, that probabilistic incoherence is defective. His vindication of standard coherence was meant to be non-pragmatic, and one of my conclusions was that a non-pragmatic vindication, along the lines he attempts, is not to be had. I criticized some of the conditions he himself laid down; they aren't all required, I argued, for a person to qualify as purely concerned with truth. Then I helped myself to standard decision theory, parts of which he meant to vindicate, squeezed all I could from the notion of purely epistemic goals, and still couldn't get the main result that he derived from his conditions. It would seem that if a non-pragmatic vindication along Joyce's lines isn't to be had even with assumptions that help themselves to the view to be vindicated, it isn't to be had at all. Epistemic rationality, I concluded, isn't to be explained as what a sheer concern with truth must endorse.

Concern with truth in one special form, though, did seem to do some explaining. That was the lesson I drew from theorems of Mark Schervish. The form is concern with truth on pragmatic grounds, but of one particular kind—or alternatively, a concern with truth that mimics such a pragmatically grounded concern. Attempted pragmatic vindications of probabilism are of course well known, and aspire to be much more general than the limited pragmatic vindication that I ventured. Whether any of these pragmatic vindications work has been widely debated, and Frank Arntzenius may be unconvinced by some of them. Nothing I showed adds anything to those debates. It does seem to be a lesson of my argument,

[1] The theorems I appealed to and the core of my argument were, as I indicated, drawn from the work of statistician Mark Schervish.

though, that any successful vindication will have to be pragmatic. More guardedly, I should say, that's the lesson unless something more can be squeezed out of notions of truth-conduciveness than I myself could identify.

Both commentators put my puzzle in ways somewhat different from what I intended. According to Arntzenius, my puzzle is that I "can see no good reason to be self-confident," no good reason not to judge my beliefs epistemically deficient. Not exactly: as Arntzenius indicates later on, whatever reasons one has for one's degrees of credence are reasons for thinking them right, and thus that one has got them right. It's just that I don't see how the reason can take a particular form: thinking—even circularly—that one's credences aim at truth in a way that is optimal given one's evidence. I thus don't see how an intrinsic concern for the truth of one's beliefs could in any way underlie epistemic rationality. (Arntzenius, as I read him, doesn't see how either, though he may think we could see the folly of such approach without any argument like mine.)

I also don't think that it is "rationally acceptable to judge one's own degrees of belief as epistemically deficient." A perfectly rational person, I would think, will not so judge. It may well be rationally acceptable, I said, to wish that one's degrees of belief were different from what they are—even when it's rational to care only about their closeness, by some standard, to full truth. Epistemic deficiency, though, is different from being unwanted. It's different even from aiming badly at truth. My puzzle brings into question not epistemic rationality but a specious way of explaining it. Can we "give a purely epistemic justification for why our belief states are as they are" (if they are ideally rational)? Arntzenius says that I think we can't, but in truth I don't know and I would hope that we could. My conclusion was that we can't give such a justification along the lines that I scrutinized, explaining epistemic rationality as somehow well aimed at the truth for its own sake.

According to Swanson, I think that epistemic rationality constrains credal value functions to be credence-eliciting. Again, not exactly: A person could be epistemically rational and value truth in all sorts of ways. He might even disvalue truth, but find himself epistemically rational against his wishes. The thesis I scrutinize in my paper, once I think I have made sense of it, allows for this. If a person is epistemically rational, goes the thesis, it is *as if* she valued truth for its own sake and could choose her credences at will. Most of us can't choose our degrees of credence at will, and a person who can't might conceivably be epistemic rational to perfection, but wish that she weren't.

One more set of preliminary remarks: Swanson suggests strengthening my characterization of concern with truth. The concern with truth, he says, should be symmetric: one should value credence in the negation of a claim, should it be false, just as one values credence in the claim should it be true. He notes that this makes no difference to the conclusions I drew, but even so, I'll register my disagreement. To be sure, the truth of S amounts to the falsehood of $\neg S$, and so trivially, the truth of S and the falsehood of $\neg S$ are of equal import. It doesn't follow, though, that the truth and falsehood of S are of equal import. Take almost any example: let S be Newtonian physics, or a value for the speed of light, or the new Hair-Brane theory in particle physics. Must uncertainty that S is true in case it is true and uncertainty that S is false in case it is false be equal failings, from the standpoint of a pure, scientific thirst for truth? I don't see why. We're comparing, say, a person who is 95 percent certain of the inverse square law for gravity when it isn't quite the correct law, with a person who is 95 percent certain that it isn't the correct law when it is. Why must their high but misplaced confidence and their correct residual doubts be of equal purely epistemic import, when its being precisely true would tell us a lot and its being not quite right would leave it wide open just what is right? I don't know which residual doubt is more important, but once we're convinced that not every 1 percent difference in the credence one might have matters equally, why think that these two do matter equally? "Credence aims at accuracy," Swanson proposes, and "a valuation of credal accuracy should not arbitrarily privilege credence in truths or credence in falsehoods by valuing them asymmetrically." I agree that a blanket policy of treating all truths one way and all falsehoods another isn't even possible, since the negation of a falsehood is a truth. I suggest, however, that a particular truth and its negation might very well be treated asymmetrically by a person who still rightly counted as valuing truth purely for its own sake.

2. THE DYNAMIC CASE: UPDATING OVER TIME

Both Arntzenius and Swanson analyze thinkers who take in new evidence over time and somehow modify their credences in its light. I agree that such an analysis is needed, and I'll turn first to Swanson's treatment. Swanson argues that a dynamic analysis changes the lesson to be drawn—at least for beings like us, with our limitations. With this I mostly disagree.

Note first a crucial feature of the static case. A coherent believer who wants only truth, recall, may wish that her credences were different from what they are. That was the central point with which I began. Note, though: in that case, if she got what she wanted, she still wouldn't be satisfied. Her credences would be different from what they are, and so her prospective valuations of the various possible arrays of credences one might have would, in this counterfactual case, be different from what they are in actuality.

Swanson's treatment of the dynamic case plays on this feature. He considers hypothetical believers "who always act to maximize expected value" and "can choose their own credences." He shows that "a believer who has non-credence eliciting value functions *and can choose her own credences* engages in some very odd doxastic behavior over time."[2] His dynamic believer, able to choose her credences anew at each updating, ends up with credences she wouldn't have wanted in the first place for the case of receiving the string of evidence that she receives.

This is quite right, as he shows conclusively. The remaining question is how it bears on the claim that epistemically rational credences do in some sense "aim at truth." For the static case, I argued, rational credences do aim at truth, but in a special way: it is as if they aimed at truth for the sake of guidance. Valuing truth in a way that mimics valuing it for the sake of guidance, though, I said, is far from the only way one could value truth for its own sake. Now the way I set up the question for the static case has a parallel for the dynamic case. It is this parallel, I'll argue, and not the case that Swanson analyzes, that bears on whether epistemically rational policies for credences and their revision can be explained as aiming at truth.

[2] Emphasis mine, and with the pronoun changed to facilitate reference.

For the static case, recall, I put the question as whether, if a person is epistemically rational in her credences, it is *as if* she valued truth and had been able to choose her credences at will. (If she valued truth in a way that made her want different credences from the ones she has, we now note, she would want not only to have those different credences, but to lose her power to set her credences at will. Otherwise she would end up, after a series of new choices of credences, with credences different from the ones she now wants.) The answer to my question depends, of course, on what qualifies as "valuing truth"—but I'll put off further discussion of that until later, and assume for now that I was right about what valuing truth in one's credences consists in. Our question now is how to pose the parallel question for the dynamic case. For the dynamic case, we suppose that the believer values truth not only for her credences at the outset, but for the credences she will come to have as new evidence crops up. What she needs to evaluate, then, is whole ways she might be disposed to form credences and update them. In actuality, we are supposing, she is epistemically rational, and so her actual epistemic dispositions, whether she wants them to be that way or not, consist in starting out with a coherent, epistemically rational array of initial credences and then updating by standard conditionalization. The question is whether she will be glad that those are her epistemic dispositions. If she in some way values truth and truth alone, will her actual epistemic dispositions be the ones she most prefers to have?

The answer to this question for the dynamic case exactly parallels the answer for the static case. What are the alternatives among which she can have preferences? As both Swanson and Arntzenius recognize, she isn't restricted to wishing to update by standard conditionalization. Swanson proposes another restriction, though, which I'll accept as an important restriction to explore. Let's confine our consideration to beings who, like us, can't keep track of their past histories of updating. Suppose, indeed, that our believer can't even aspire to more, that she is constrained to wish only for epistemic dispositions that don't require keeping track of such matters as how many times she has updated. On each updating, we require for the world as she wishes it were, she must apply a rule that takes her current credences and the new item of evidence, and on the basis of these alone delivers a revised array of credences. What

dispositions, under this restriction, will she most prefer to have? That is our question.

Swanson provides the machinery that delievers an answer to this question. Take the "report relation" R that Swanson defines, which takes actual to wished-for credences. Look, as he shows that we must, for an array of dispositions that mimic updating from her actual credences by standard conditionalization and then "applying R at the end." Because of the informational restriction, we must now, I agree, further require that her way of valuing truth yields a report relation that is injective (that is, that it is a one-to-one function from the interval [0,1] onto itself). As he notes, however, this isn't a severe restriction; it allows for many report relations that aren't the identity relation—that aren't the R of a believer who most prefers the epistemic dispositions that she in fact has.

Here are the dispositions she most prefers to have (though so long as R isn't identity, she doesn't in fact have them): the dispositions are, in effect, at each stage as new evidence arrives, to revert to her epistemically rational credences, apply standard conditionalization, and then go to the new credences that, in actuality, she prefers for the case of having that evidence. This works as follows. Let ρ_0 be her actual, epistemically rational credences at time 0, and for discrete times $t = 1, 2, \ldots$, let ρ_t be the credences that, with her actual dispositions, she would have at time t having received a string of evidence $E_1, E_2, \ldots, E_t$. What arrays of credence $\sigma_0, \sigma_1, \ldots, \sigma_t$, we now ask, does she wish she were disposed to have on receiving that string of evidence. She wishes, as Swanson says, that each σ_t were the one she would get by starting out with her actual initial credences ρ_0, updating by standard conditionalization, and applying R at the end. But a non-standard updating rule that she can wish for would accomplish just that. (Indeed it is a rule that Swanson considers, though it doesn't work for the situation that Swanson considers, where the believer is stuck having to wish for states where she could wish further and get what she then wished.) Let her wished-for initial credences σ_0 be the ones that result from applying R to her actual initial credences ρ_0. Let her wished-for dispositions to update be this: that on receiving each new piece of evidence, she update as if she first had reverted to the credences ρ_{t-1} that she is actually disposed to have, then had updated these by standard conditionalization, and finally had applied the report relation R to the result.

This gives her a wished-for updating rule that fits Swanson's restriction on wished-for information. The rule, more fully put, consists in first (i) applying to her wished-for credences σ_{t-1} the inverse R^{-1} of the report relation, yielding her rational credences ρ_{t-1}, then next (ii) applying standard conditionalization C_t, defined as $C_t(\rho_{t-1}) = \rho_{t-1}(\cdot/E_t) = \rho_t$, and finally, (iii) applying the report relation to the result to get $\sigma_t = R(\rho_t)$. Her wished-for updating function is thus RC_tR^{-1}, the transformation that results from applying successively the transformations R^{-1}, C_t, and R. This may be messy, but applying this updating rule would, with enough sheer calculating power, require only keeping track of one's current credences and what the new evidence is.

This dynamic parallel to the static case differs sharply from the case that Swanson analyzes. I examine only what the rational believer who values truth actually wants. Swanson examines a case where the believer, on the arrival of each new piece of evidence, gets what she wants and so forms new preferences which are then accorded at the next updating. This, as he shows, isn't something to want—unless one wants precisely the initial credences one has. His treatment plays, as I have said, on a feature that the dynamic and the static cases share: that in case the believer isn't satisfied with her credences, if she got what she wants she still wouldn't be satisfied.

Would this feature itself, though, indicate that she doesn't genuinely want truth in her credences? Does it show that she fails really to value truth and truth alone? If it does, then perhaps the dictum that belief aims at truth can still be interpreted as correct. We can still maintain that any rational believer who values truth *genuinely* will be glad she has the credences she does.

But this feature indicates no such thing. All sorts of things we might genuinely value in beliefs will display this feature. The suicide prefers self-inflicted death to his prospects otherwise—but once he kills himself, he no longer has this preference. His preference is none the less genuine. Or take an instance that is more complex: I want comfort, but I also want to be emotionally braced for rude surprises. I want not to be completely terrified all the time, but still to be somewhat prepared for the things I dread. What credences would, on balance, prospectively best meet these and my other competing desiderata? They may not be the credences I actually

have and that I regard as epistemically rational. Perhaps, for the sake of comfort, they'd discount the likelihood of some of the things I fear—but still not too much, or I'll be too unprepared if terrifying things do happen. What credences I most want to have will thus depend, among other things, on how likely I now take various nasty eventualities to be. For that reason, if I had the credences I actually most want, the calculations I now make would no longer apply. I'd want even lower credence in fearsome things that might befall me. None of this means, though, that I don't now genuinely value comfort as a benefit that my credences might yield.

I conclude, then, that the dynamic case works like the static one—with a qualification. A being fully coherent in belief and preference might intrinsically value truth and truth alone and still want a credal policy different from her actual ones. In the dynamic case, she might want both different initial credences and a different updating rule. As Swanson indicates, the updating rule she wants will in some cases demand extraordinary amounts of information. Not so, however, in cases where her epistemic preferences yield, in Swanson's terms, a report relation that is injective. Then, the rule she most wants can run on the same information as standard conditionalization: one's credences prior to the new evidence and what the new evidence is. Valuing truth, then, even in this restricted way, needn't lead an ideally rational person to want the credences she has. Epistemically rational credences, then, can't be explained just as being what you'd want if you valued truth and truth alone.

Arntzenius, for the dynamic case, starts out with just the right question. "What should I now regard as epistemically the best policy for updating my degrees of belief in light of the evidence I will get." He shows, for the particular case he considers, that the policy will depart from standard conditionalization as its updating rule. I agree, as I have indicated in my treatment of Swanson. He finds problems with this, however. First, it goes against diachronic Dutch book arguments, and if we lose Dutch book arguments, we have no answer to why credences ought to satisfy the axioms of probability—why, as I'm using the term, they ought to be coherent. Dutch book arguments, though, are pragmatic, not purely epistemic, and I haven't questioned pragmatic arguments for classical decision theory. My point is that we can't get a certain kind of purely epistemic argument to work. As for why to have degrees

of belief that satisfy the axioms of probability, that is an excellent question, but not one that I took up. I considered only degrees of belief in a single proposition.

Arntzenius's second problem with dropping standard conditionalization is that one loses "the ability to set one's degrees of belief so as to maximize the current expected epistemic utility of those future degrees of belief." Here what I said about Swanson applies. In the linear case, the one that Arntzenius chiefly analyzes, Swanson's report relation R isn't injective. We can still ask Arntzenius's question of what, by my actual lights, would be my prospectively best updating policy. The policy that looks prospectively best by my initial lights will still look prospectively best over time as new evidence comes in. But the policy will make heavy informational demands; it can't prescribe credences as a function just of what one's credences are before a piece of evidence comes in and what that evidence is. Arntzenius may be suggesting this when he says that if I had my desired credences, "I would lose the information as to what I should do were I to learn $\neg E$" (section 2). I need lose it, though, only in the sense that the information won't be given by my desired credences. Conceivably I might have the information in some other form. One form the information might take is in the double bookkeeping that Arntzenius proposes, having as one's information both one's "epistemic" and one's "prudential" utilities. If, on the other hand, the Swanson report relation R is one-to-one, the needed updating rule will require only the information that standard conditionalization requires.

Arntzenius draws the lesson, "if one's epistemic utilities are linear, then maximizing the expected epistemic utility (by one's current lights) of one's degrees of belief can make it impossible to maximize the expected epistemic utility (by one's current lights) of one's degrees of belief at a future time" (section 2). He himself, though, goes on to propose a way out, and it is important to bear in mind two qualifications to what I just quoted. First, we can imagine updating in a way that achieves both these goals if the policy can draw on enough information, as with Arntzenius's own proposal of keeping double books. Second, some perverse cases differ from the linear one that Arntzenius is treating, in that the Swanson report relation R is injective. For these cases, we don't face this dilemma.

I mostly agree with Arntzenius about his suggested way out, his proposal of keeping two books with two different arrays of credences. An agent who acts as well as believes will need "prudential" credences anyway, to guide her actions in pursuit of new evidence. The Schervish result shows that, purely for guidance, the rational agent will want the credences she has. If she also values some form of closeness to truth in her epistemic credences, just for its own sake, she might indeed then wish she kept such double books, with one array of credences to guide her and another to maximize closeness to truth by the standards she embraces. She might wish this, Arntzenius shows, even if she has no other goal than closeness to truth on some specification.

I agree with Arntzenius too that such a wish is ridiculous. First, of course, it will satisfy the believer's preferences only if she cares intrinsically solely about her "epistemic" credences and not about her guiding "prudential" ones. Otherwise, she'll have to find some array of guiding credences that best answer a balance of competing demands: the demand to govern her assessments of expected epistemic utility, and the demand of being truthful in the way she values intrinsically. (Like things would go for wanting credences that will comfort one, enhance one's social dominance, stave off depression and anxiety, and the like. The best thing might be to keep one's epistemically rational credences for purposes of guidance, and have a separate set of cuddly or enlivening credences for these "side" purposes.) Second, if she had the "epistemic credences" she wishes for, they would be idle.

One interesting lesson that Arntzenius draws is worth stressing. He has given, he says, "a purely epistemic argument for updating one's prudential degrees of belief by conditionalization, on the grounds that such updating guarantees cross-time consistency of epistemic utility maximization" (section 2). Even if one's goals are purely epistemic, he shows, epistemically rational credences can offer prospectively optimal guidance in achieving those goals. They can do so not only by guiding action in pursuit of new evidence, but by guiding assessments of possible epistemic states for their prospective closeness to truth by some standard. In these senses, we can have a purely epistemic vindication of epistemic rationality.

3. EPISTEMIC UTILITIES AND COHERENCE

Arntzenius in section 3 questions the whole notion of epistemic utilities. I should be happy with such questioning: the lesson I drew was a debunking one. Whether or not talk of epistemic utilities makes sense, I argued, no such utilities play any role in explaining epistemic rationality. (I would now admit an exception to this, namely the roles epistemic utilities played in the last paragraph above.) What might play such a role, I said, is rather a tie to mundane, non-epistemic utilities—to the utility of happiness, wealth, health, or some other such things. I admit I can't myself shake off a residual sense that a pure concern for truth is intelligible and might sometimes be reasonable. Nothing in my debunking, though, required making precise sense of the line I found wanting.

Arntzenius imagines an immobilized robot Hal, and has me asking, "Suppose you just wanted Hal's current degrees of belief to be accurate, what degrees of belief would you give him?" That depends on what I mean by "accurate," he responds—my point exactly. "Gibbard is asking an unclear question." Yes, but as Arntzenius goes on to recognize, I was asking questions like this in order to expose them as unclear. According to Arntzenius, though, I still think the question to be well-defined, though with only person-specific answers. I wouldn't put it that way, and I'm not clear just what such a thought would amount to. My point was that this ill-defined question suggests a whole family of well-defined questions. Tell us just what you mean by "accurate" and you will have indicated a particular question in this family.

Why then have degrees of belief? A big question, this, which I didn't vaunt myself as able to answer. As I think Arntzenius sees, he and I are pretty much in accord on this. "When one's only goal is truth why should one's epistemic state satisfy the axioms of probability?" To this I offered no answer. In the first place, I considered credence just in a single proposition, and so most of those axioms didn't come into play. In the second place, my aim was to refute a certain kind of purely epistemic vindication of standard coherence, and unless some replacement is found, that leaves only the familiar sorts of pragmatic vindications: Dutch book arguments and more comprehensive representation theorem arguments. I may

be more optimistic about representation theorems than Arntzenius is, but that's another story, and his expertise on such matters far exceeds mine.

"Why think a rational person must have purely epistemic preferences over all possible belief distributions?" There's no reason—or at least no reason they can't all be zero—unless intrinsic curiosity is itself a requirement of reason. If it is, then the fully rational person is prone act, in some conceivable circumstances, just to find something out, for no further reason. Having learned from Arntzenius of the Hair-Brane theory, I'm curious, and given the opportunity, I might expend resources and effort to garner evidence of its truth or falsehood. Does this require a full set of utilities over my possible states of belief? The story here would be the same as with the rest of decision theory. On the one hand, I can cross bridges when I come to them, and form no preferences until I need them. If, though, I go to an extreme of looking before I might leap, deciding in advance every decision problem that is even conceivable, then consistency may require fully determinate utilities for everything.

If I do have well-defined utilities for everything, can we separate out a purely epistemic component of those utilities? I don't know. My own question was a hypothetical one about a being whose *sole* intrinsic concerns are with her degrees of belief. The being, I supposed, is ideally coherent in her credences. Such a being, I now agree, will still need epistemically rational credences for purposes of guidance. Only epistemically rational credences, after all, will be prospectively optimal, by the being's own lights, as guides in seeking out evidence or assessing the value of possible states of credence. If, though, the being is passive, with nothing she can do but sit back and await new evidence, then thirst for truth as such can't explain her epistemic rationality. Epistemically rational credence can't be explained just as aimed at truth.

4. THE OTHER PUZZLE

What, then, of guidance value? The two commentaries focused on the negative thesis of the paper, on the puzzle, if I am right, that aiming at truth as such can't underlie epistemic rationality. The Schervish results lead, though, to another puzzle. Does guidance value somehow underlie the nature of epistemic rationality? I haven't yet seen to the bottom of all this, and I need help.

The main Schervish result is striking: epistemic rationality is what a fully coherent person will want if she is concerned with her epistemic states solely as guides. Epistemic rationality isn't everything one could want from one's beliefs: one can want comfort, or self-affirmation, or any of a host of other things, and one can want truths just for the sake of having them. Guidance value is just one component of the value that one's beliefs may have. Schervish, though, demonstrates a tight relation between guidance value and epistemic rationality, and it would be strange if the nature of epistemic rationality has nothing to do with this striking relation. But although it is *as if* an epistemically rational person had chosen her credences for the sake of guidance, of course she didn't. She couldn't indeed have conducted a full, rational analysis of prospective guidance values without epistemically rational credences already in place. Exactly what, then, if anything, *is* the bearing of the Schervish findings on the nature of epistemic rationality? That is a second puzzle.