Bridging Rationality and Accuracy[1]

Forthcoming in the Journal of Philosophy

Rationality and accuracy have to be connected somehow. It's not a coincidence that rational people tend to do a better job forming accurate beliefs and navigating the world than irrational people. But *how* are they connected? The aim of this paper is to explore this question.

The thought that some connection must exist between rationality and accuracy is certainly not a new one. Stewart Cohen notes that: "A variety of philosophers from Descartes to the present have presupposed the view that justification and truth are conceptually related – that there is an internal connection between a belief being justified and being true."[2] But, he later concludes: "…it has turned out to be a difficult matter to say precisely what the connection comes to."[3]

Although much of the debate about "the truth connection" has focused on its relevance to internalist and externalist theories of epistemology, two more recent movements in epistemology motivate serious reconsideration of the connection between rationality and accuracy. The first is the "accuracy first epistemology" project – a project aimed at deriving rational requirements from accuracy based considerations. The second is the higher order evidence debate, which is concerned with how we ought to respond to evidence about our own rationality.

In this paper I will outline two ways of thinking about the connection between rationality and accuracy and I will argue that what we say about higher order evidence will depend on how we think about the nature of this connection. The higher order evidence case demonstrates a more general lesson for accuracy-first epistemology: which first order views the accuracy-based considerations motivate depends on what we think the connection between rationality and accuracy amounts to.

[2] Cohen, S. (1984) "Justification and Truth" *Philosophical Studies* 46(3): 279-295, p.279.
[3] Cohen *ibid* p.292-3.

The paper is structured around three questions:

Preference Question #1: Why should we prefer rational credences to irrational credences? (This question is similar to what Sophie Horowitz[4] has called "the value question").

Preference Question #2: Why should we prefer credences rationalized by our total evidence to credences rationalized by a proper subset of our evidence?

Constitutive Question: What constitutive connection is there, if any, between rationality and accuracy?

These questions will be made more precise as we proceed.

For the sake of simplicity, I will make three assumptions: first, *precision* – the claim that (rational) agents' doxastic states are representable by precise credence functions; second, *uniqueness* – the claim that there exists a unique function from evidence to doxastic states that assigns to each body of evidence the doxastic state that is rational given that evidence; and third, *introspection* – the claim that rational agents are always certain about what their evidence is. I do not endorse all of these assumptions,[5] and although I think that (a version of) the arguments I develop in this paper are available to those who reject these claims, I will not argue for this here. I will be satisfied if I can provide answers to my three questions that apply, at very least, in cases in which the rational response to evidence is unique and precise, and when agents are certain about what their evidence is.

Here's the plan: In section 1, I make the preference questions more precise. In section 2, I introduce one picture of how rationality and accuracy are connected. This picture, I will show, is presupposed in many accuracy-based arguments for rational requirements. Roughly, it is one according to which the correct theory of rationality is the *epistemic plan* that one should regard, a priori, as the best plan (from an accuracy standpoint) to follow. In section 3, I show why this picture is unavailable if one accepts certain commonly held views about higher order evidence (what are sometimes called "conciliatory" or "calibrationist" views). The remainder of the paper is devoted to

---

[4] Horowitz, S. (2013). "Immoderately Rational." *Philosophical Studies* 167(1): 1-16.

[5] Indeed I have argued in print against both *precision* (as applied to cognitively limited agents) and *uniqueness*. (See my "Chilling out on Epistemic Rationality: A Defense of Imprecise Credences (and other imprecise doxastic attitudes)," *Philosophical Studies* (2012) 158:197-219 and "Permission to Believe: Why Permissivism is True and What it Tells Us About Irrelevant Influences on Belief," *Noûs* (2014) 48(2):193-218)).

developing an alternative way of thinking about the connection between rationality and accuracy, which is not only consistent with, but *motivates* calibrationist views about higher order evidence. Roughly, the alternative account will be one according to which the correct theory of rationality is the epistemic plan that one should regard, a priori, as the best plan (from an accuracy standpoint) *to make*.

1. Refining The Preference Questions

The goal of this section is to make the preference questions more precise. To begin, I need to introduce some terminology.

(a) Epistemic Value (Accuracy)

An epistemic value function (or scoring rule), A, is a function that takes as input a credence function, *c*, and a state of the world, s, and outputs a number, A(*c*,s) representing how accurate the credence function is in that state.

(b) Expected Epistemic Value (Expected Accuracy)

The *expected* accuracy of a credence function *c* is calculated relative to an epistemic value function A, a probability function *p*, and a partition of states S, as follows:

$$EA_{A,p}(c) = \sum_{s \in S} p(s)A(c, s)$$

In other words, the expected accuracy that *p* assigns to some credence function *c* is the sum of the values *c* would have in states s, weighted by the probability that those states obtain.

(c) Credence Functions Under Descriptions

Sometimes we're interested in the expected accuracy of a credence function *under some description*. For example, I might want to calculate the expected accuracy of *Jane's credence function* or *the credence function I will have tomorrow* even if I don't know which credence functions satisfy these descriptions. The expected accuracy of a credence function under description D is:

$$EA_{A,p}(D) = \sum_{s \in S_d} p(s)A(D(s), s)$$

where S_d is the set of states in which some credence function satisfies D, and D(s) is the credence function that satisfies D in s.

In what follows the distinction between the expected accuracy of credence functions *specified as such* and credence functions specified *under a description* will be important. When I talk about the expected accuracy of a credence function *specified as such* I am assuming that the agent is aware of the identity of the credence function in question. (That is, the agent is aware that the credence function assigns, for example, credence 0.1 to p, credence 0.72 to q, and so forth). Lower case italicized letters are used as stand-ins for credence functions specified as such. When I talk about the expected accuracy of a credence function *under a description*, I leave it open whether or not the agent is aware of the identity of the credence function in question.

(d) Strictly Proper Value Functions

There are many different epistemic value functions, and my arguments do not rely on the use of any particular one. All I assume is that the epistemic value function is *strictly proper*. Strictly proper value functions are ones according to which every probability function regards itself as more expectedly accurate than any other function *specified as such*. This means that any probability function, *f*, will assign greater expected accuracy to itself than to a distinct function *g*. (However, it may regard as more expectedly accurate some other function *specified under a description* such as "the probability function that assigns 1 to all truths and 0 to all falsehoods.")[6]

---

[6] For a discussion of the motivation for using strictly proper scoring rules see Gibbard, A. (2008) "Rational Credence and the Value of Truth" *Oxford Studies in Epistemology Volume 2*, Oxford University Press; Greaves, H. and Wallace, D. (2006) "Justifying conditionalisation: conditionalization maximizes expected epistemic utility" *Mind* 115(459): 607-632; Horowitz, *op. cit.*; Joyce, J. (2009) "Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief" in F. Huber & C. Schmidt-Petri (eds.), *Degrees of Belief*. Synthese Library. 263-29; Moss, S. (2011) "Scoring Rules and Epistemic Compromise" *Mind* 120 (480):1053-1069 and Pettigrew, R. (forthcoming) *Accuracy and the Laws of Credence*, Oxford University Press.

I will now describe two principles aimed at making our judgments about the sense in which we should prefer the rational to the irrational, and our total evidence to a subset of that evidence, more precise. The preference questions then become questions about why the two principles I describe are true (if, in fact, they are).

The first principle concerns the way in which we should prefer rational credences to irrational ones. I will set the stage by describing the case of Bill, a competitor on a game show.

GAME SHOW 1

Bill is on a game show in which contestants make bets about tomorrow's weather. Bill's total body of evidence is E. Bill hasn't being doing so well, so the host offers him a special deal: "In this round, you will be betting on whether or not it will rain tomorrow (RAIN). I will give you a choice between a number of different pills. If you take the rationality-pill, you will have the credence in RAIN that is rational given your evidence.[7] Alternatively, you can choose any number between 0 and 1 that you know is an *irrational* credence to have in RAIN given your evidence. I will then give you an irrationality-pill that will make you have the credence corresponding to your choice."

Which of these two options should Bill choose? Plausibly, if Bill is trying to do well on the game show, he should pick the rationality pill.

I will use "R(E)", throughout, as a stand-in for the description: "the rational credences given E". It is important to remember that, since R(E) is standing in for a *description,* when I talk about an agent's attitudes towards R(E), I do not assume that the agent knows the identity of R(E).

The judgment about Bill suggests:

---

[7] Here, and throughout, I am talking about rationality in the sense of *propositional* rather than *doxastic* justification, so there is no problem with gaining a rational credence, in this sense, by taking a pill.

> *RatPref* (preference)*:* For any body of evidence, E, if E is your total evidence, you should prefer R(E) to any function *i* which you know differs from R(E).[8]

Prefer in what sense? At very least, the sense that is relevant to your interest in doing well on game shows. More precisely, I will follow Allan Gibbard[9] in using "the guidance value" of a credence function to refer to the value of that credence function as a guide to choice in pursuit of other values. Gibbard, employing a result from Schervish,[10] shows that the credences with the highest guidance value are the credences that maximize expected *accuracy* according to a strictly proper value function.[11] So we can also think of RatPref as the claim that you should assign greater expected accuracy (as measured by a strictly proper scoring rule) to R(E) than to *i* when you know that *i* differs from R(E):

> *RatPref* (expected accuracy): For any body of evidence E, and for any *r* such that
> R(E) = *r*,
>
> $$EA_{A,r}(R(E)) > EA_{A,r}(i)$$
>
> Where $r(R(E) = i) = 0$.

RatPref (expected accuracy) says that if the rational credence function given E is *r*, then *r* will assign greater expected accuracy to the credence function *under the description* "the rational credences given E" than to any credence function *specified as such* which *r* is certain is irrational. Since the preference version and the expected accuracy version are so tightly connected, I will simply talk about "RatPref," sometimes in terms of preferences and sometimes in terms of expected accuracy.

---

[8] Recall that I am assuming, throughout, that agents know what their evidence is. If this weren't the case, it might be strange to think that a rational agent should give preference to R(E), when she doesn't know that E is *her* evidence.

[9] Gibbard, *op. cit.*

[10] Schervish, M. (1989) "A General Method for Comparing Probability Assessors," *Annals of Statistics* 17(4):1856-1879.

[11] This is not to say that it can never be in your best interest to have inaccurate doxastic states. Sometimes, for example, having a belief with a certain *content* will promote your aims, regardless of its accuracy. The guidance value of a credence function is determined by the value of that credence function as a guide to other pursuits, insofar as your success depends on the accuracy of your doxastic states, and not their content.

I find a principle like RatPref attractive, and closely connected to whatever constitutive connection exists between rationality and accuracy. But what I'm interested in here is *what exactly it is about rationality*, and its connection to accuracy, that could explain why we should *always* expect the rational credences, whatever they are, to do better than some irrational credence function specified as such. After all, with any other generally reliable but sometimes fallible indicator of truth, it can sometimes be reasonable to expect that the indicator is going astray. I might think, for example, that my thermometer is generally reliable, but that, *in this particular case*, whatever it says is likely to be less accurate than some other temperature, which I know differs from its reading. (For example, maybe I know that the thermometer does not operate well at certain altitudes). So why is it that, when it comes to rationality, a generally good but fallible guide to the truth, I can *never* assign less expected accuracy to the rational attitude (whatever it is) than to some particular attitude that I know is irrational? The thought that we must *always* expect the rational credence given our evidence to be best is indicative of an especially strong sense in which we take rationality to be a good guide to truth.

Let's now move on to the question of why one should be especially interested in the credences that are rationalized by one's *total* evidence. What would be wrong with preferring a credence that only a subset of one's evidence supports? Consider:

GAME SHOW 2

On an advanced round of the game show Bill is offered a slightly different deal. As before, he can choose a pill that will give him the rational credence in RAIN, given his total evidence E. Alternatively, he can choose a particular *subset of his evidence*, which he knows supports a credence that differs from the credence that's rational given his total evidence. The host will then give him a pill that will cause him to adopt the rational credence given his chosen subset.

It seems like Bill should choose the pill that will give him the credence rationalized by his *total* evidence. This suggests:

> *TotPref* (preference): For any body of evidence E, and any E- that is a particular set of propositions that is a proper subset of E, if E is your total evidence, you should prefer R(E) to R(E-), when you know that R(E) differs from R(E-).[12]

Once again, the relevant sense of "prefer" is the sense that is relevant to making good choices in pursuit of other values. So we can also think of the principle as follows:

> *TotPref* (expected accuracy): For any body of evidence E, and for any r such that R(E) = r:
> $$EA_{A,r}(R(E)) > EA_{A,r}(R(E-))$$
> where E- is any particular proper subset of E and $r((R(E) = R(E-)) = 0$

To see how RatPref and TotPref can come apart, imagine that an agent doesn't know what credence function a body of evidence E-, which is a proper subset of her total evidence E, supports. RatPref will tell the agent to prefer R(E) to any credence function *specified as such* that the agent knows is irrational. However, RatPref doesn't say anything about whether the agent should prefer R(E) to credences that the agent knows are irrational, but that are being considered *under some description*. For example, RatPref doesn't forbid the agent from preferring the credence function satisfying the description "OMN: the credence function that assigns 1 to all truths and 0 to all falsehoods" to R(E), even though the agent knows that OMN is irrational. For the same reasons, if the agent doesn't know which credence function satisfies R(E-), then, even if the agent knows that R(E-) is irrational, RatPref alone doesn't tell the agent to prefer R(E) to R(E-).

RatPref and TotPref make explicit the way in which agents should prefer rational credences to irrational ones, and credences supported by their total evidence to credences supported by a proper subset. Before proceeding, I want to be clear about what these principles are *not* saying. First, RatPref and TotPref are not meant to be addressing the

---

[12] Note that, since E- refers to a *particular* proper subset (or a proper subset of evidence *specified as such*), E- can't be, for example, "the proper subset of my evidence that excludes all misleading evidence." Also, recall, once again, that we are assuming that agents know what their evidence is.

question: "Why be rational?" or "Why care about rationality?"[13] RatPref and TotPref are principles about what you *should* prefer, and the "should" here is a rational should. The question is: why is it *rational* to prefer, for the purposes of doing well on the game show, or being accurate relative to a strictly proper value function, the credence that is rational? If you're skeptical about the value of rationality, the claim that it's rational to prefer the rational credences may not have much force. But the preference questions are by no means trivial either. I am not asking why it's rational to be rational. The fact that one should *have* the credences that are rational doesn't mean that one should *prefer to have* the credences that are rational. For example, it's rational for Bill to prefer to have the credence function that assigns 1 to the truth about RAIN even though it's not rational for Bill to *have* a credence function that assigns 1 to the truth about RAIN. So the question of whether one should prefer, for accuracy purposes, to have the credences that one, in fact, should have is a substantive question.

2. Best Plan to Follow

In this section I describe one picture of how rationality and accuracy are connected that will allow us to answer the three questions I posed. This picture, we shall see, is implicit in much of the "accuracy-first" epistemology literature.

To begin, I need to do a bit of setup. Call a function from potential bodies of evidence to doxastic states, an "epistemic plan." (This sort of function is what Greaves and Wallace[14] call an "available act"). Now imagine that a perfectly rational agent with no empirical information (I'll call her "Pria" for (**p**erfectly **r**ational **i**gnorant **a**gent)) is contemplating what plan she would "program" into a robot if her goal was to maximize the expected accuracy of the credences the robot will have upon entering the world. By "program" I mean that Pria is choosing a function that will *fully determine* the credences the robot will have, as a function of the evidence that it receives.

More precisely, suppose that Pria is considering a partition of states s in which the robot receives some body of evidence (we'll call this set of states "S"). The partition must

---

[13] These questions are discussed by Kolodny, N. (2005) "Why be Rational?" *Mind* 114(455):509-563 and Broome, J. (2007) "Wide or Narrow Scope." *Mind* 116(462):359-370.
[14] Greaves and Wallace, *op. cit.*

be sufficiently fine grained so as to fix, for any s ∈ S, (a) the evidence that the robot receives in s and (b) the accuracy score of any credence function that the robot might adopt in s. Let $\mathcal{P}$(s) be the credence function that a plan $\mathcal{P}$ recommends given the evidence that the robot receives in state s. If Pria's probability function is *Pr,* then Pria is looking for a plan that maximizes the following quantity, which we'll call the "expected accuracy of $\mathcal{P}$":

$$EA_{A,Pr}(\mathcal{P}) = \sum_{s \in S} Pr(s) * A(\mathcal{P}(s), s)$$

In other words, the expected accuracy of a plan is the weighted average of the accuracy scores one would get from adopting the credences recommended by the plan. Call the plan that maximizes the quantity above the *accuracy-optimizing epistemic plan* or $\mathcal{A}$ for short. We can also talk about the function from evidence to belief states that assigns to each body of evidence the doxastic state that is *rational* given that evidence. Let's call that *the rational epistemic plan* or $\mathcal{R}$ for short. One way of answering the constitutive question, then, is to claim that $\mathcal{R} = \mathcal{A}$. In more detail:

> **Best-Plan-to-Follow:** The rational epistemic plan is the one that a rational agent would choose, a priori, if she were aiming to maximize the expected accuracy of the credences that an agent following the plan would adopt.[15]
> The rational credences given a body of evidence are those that the rational epistemic plan recommends given that body of evidence.

Let me make three quick notes about Best-Plan-to-Follow before discussing it's plausibility: First, note that this way of answering the constitutive question does not give us a *reductive* account of rationality, since it says that the rational epistemic plan is the one that *a rational* agent would regard, a priori, as accuracy optimizing. Second, note that "assign 1 to all and only the truths" is *not* the accuracy-optimizing plan, since an epistemic plan is a function from bodies of *evidence* to credence functions, and Pria won't know a priori which epistemic plans will result in the robot assigning 1 to all and only the truths in the actual

---

[15] By "follow" I don't mean anything stronger than "conform to."

world. Finally, note that there is a sense in which this picture of epistemic rationality is essentially diachronic. It says, in effect, that what it is rational to believe given a body of evidence is what you'd want yourself to believe (if you were rational) *prior* to getting the evidence.

Best-Plan-to-Follow is *prima facie* plausible. For if it were false, and if Pria knew this, she could think to herself as follows: "Here are two plans, the rational one, and the accuracy-optimizing one. They are different. So I hope I don't go into the world and form credences rationally! If I want to be accurate, I'd rather adopt these particular credences (which I know are irrational) in response to E, than adopt these other particular credences (which I know are rational) in response to E!" It would certainly be surprising if this were a rational stance to take a priori.

Best-Plan-to-Follow is also in the background of much of the work being done in "accuracy-first" epistemology. For example, Greaves and Wallace[16] argue for the claim that "conditionalization maximizes expected accuracy." They show that *prior to getting evidence,* one should want to revise one's credences by updating by conditionalization. If one wants to use the Greaves and Wallace result to motivate the claim that, *upon gaining evidence*, it's rational to have the credences that result from conditionalizing on a rational credence function, one needs some principle about how facts concerning the expected accuracy of plans *prior* to receiving evidence bear on the rationality of belief states *upon receiving evidence*. Best-Plan-to-Follow is exactly the principle that fills this gap. It says that the rational credences given any body of evidence are the ones that the initial rational credence function (that is, Pria's function) regards as accuracy optimizing. We know from Greaves and Wallace that Pria will regard conditionalizing on the evidence as maximizing expected accuracy. Thus, it follows from the Greaves and Wallace result, in combination with Best-Plan-to-Follow, that the *rational* credences given any body of evidence are those that result from conditionalizing the rational prior probability function on the evidence received.[17]

---

[16] Greaves and Wallace, *op. cit.* Subsequent references to this work are given in the text.

[17] That's actually not quite right. I show in Schoenfield, M. (ms.) "Conditionalization does not (in general) maximize expected accuracy" that the Greaves and Wallace result only obtains if, for any proposition E that an agent thinks might constitute her (or her robot's) total evidence in the future, the agent is antecedently certain that E will be true if and only if E is her total evidence. The update procedure that maximizes expected

Best-Plan-to-Follow plays an important role in a number of other accuracy-based arguments for principles of rationality, since such arguments frequently take the form of showing that the *initial* rational credence function has certain features. For example, in defending the Principal Principle, Richard Pettigrew writes: "I will talk throughout of norms that govern an agent's initial or ur-credence function: that is, the agent's credence function at the beginning of his or her epistemic life, prior to accumulating any evidence." [18] But suppose we want to make a judgment about the rational response for someone *with* some evidence – say someone whose only evidence is the proposition that the chance of p is *x*. Pettigrew's argument will only support the claim that it is rational to assign credence *x* to p if Best-Plan-to-Follow is assumed. So, not only does the account have some prima facie plausibility, it underlies much of the project involving the derivation of rational requirements from accuracy-based considerations.

Let's move on to the Preference Questions. I will now show that if we accept Best-Plan-to-Follow, we will be able to give an argument for RatPref and TotPref so long as we assume that *Pria is certain that her plan* ($\mathcal{A}$) *is the rational plan*.

Here's the argument: Assume Pria is certain that her plan is the rational plan. If she is aiming to maximize the expected accuracy of her robot, she will impart this certainty onto it. After all, she will think: "If the robot assigns credence 1 to any proposition of the form: "the rational credence function given evidence E is *r*" that I am certain of, the robot is guaranteed to have credence 1 in a truth!" So Best-Plan-to-Follow, in combination with the assumption that *Pria* is certain of which plan is rational entails that:

> *Rational Omniscience*: For any body of evidence E, every rational agent will always be certain of the proposition "the rational credence function given E is *r*" where *r* is rational given E.

---

accuracy is general (even when this condition is not satisfied) is what I have called "conditionalization*." However, the differences between conditionalization and conditionalization* don't matter for the purposes of this paper.

[18] Pettigrew, R. (2012). "Accuracy, Chance and the Principal Principle." *Philosophical Review* 121(2):241-275, p.243.

I do not want to evaluate the plausibility of *Rational Omniscience* at this point.[19] My aim, in the remainder of this section, is just to show that RatPref and TotPref can be derived from *Rational Omniscience*. In the next section, I'll discuss an alternative way of thinking about the connection between rationality and accuracy that doesn't involve this commitment.   But, as we'll see, according to the sorts of views that motivate rejecting *Rational Omniscience*, TotPref is false, and RatPref is questionable. So, for better or for worse, our sympathies for principles like RatPref and TotPref are closely connected to the thought that rational agents are certain of what is rational.

Suppose that *Rational Omniscience* is true and so I, a rational agent, know that my evidence, E, supports *r*. What could be wrong with thinking that, in this particular case, the expected accuracy of the credence function supported by my evidence is less than the expected accuracy of some other credence function, *i*? Here's the argument (which is inspired by an argument from Sophie Horowitz[20]): Suppose for reductio that it could be rational for me to think: "My evidence supports *r*, but *i* is more expectedly accurate than the credence function that my evidence supports." Then, since I am certain that the evidence supports *r*, I must think that *i* is more expectedly accurate than *r*. But if it's rational to assign greater expected accuracy to *i* than to *r*, then it can't be rational to adopt *r*! This is intuitive, and this intuition is captured by the fact that we're using strictly proper scoring rules, which means that *r* must assign greater expected accuracy to itself than to any other credence function specified as such. Thus, if it's rational to have a credence function that assigns greater expected accuracy to *i* than to *r*, the rational credence function can't *itself* be *r*, contrary to our stipulation.

A similar line of reasoning supports TotPref. Suppose that *Rational Omniscience* is true and so I, a rational agent, know that my evidence, E, supports *r*, and that some particular subset of my evidence, E-, supports *i.* What could be wrong with thinking that, in this particular case, the expected accuracy of the credence function supported by my total evidence E is less than the expected accuracy of the credence function that E- supports? Here's the argument: Suppose for reductio that it could be rational for me to think: "The

---

[19] Though see Titelbaum, M. (forthcoming) "Rationality's Fixed Point" *Oxford Studies in Epistemology Vol. 5*. Oxford University Press, for a defense.
[20] Horowitz *op. cit.*

credence function that the subset of my evidence, E-, supports is more expectedly accurate than the credence function that my total evidence supports." Then, since I know that E supports *r*, and E- supports *i,* I must think that *i* is more expectedly accurate than *r*. But, for the reasons given above, if it's rational to have a credence function that assigns greater expected accuracy to *i* than to *r*, the rational credence function can't itself be *r*, contrary to our stipulation.

You might have the following worry about these arguments for RatPref and TotPref: when we judged that Bill should choose the rationality pill, or the total evidence pill, we thought: *even if Bill doesn't know* which credences are rational, he should prefer the rationality pill. And *even if Bill doesn't know* which credences some proper subset of his evidence rationalizes, he should prefer the credence that's rational given his *total evidence*. Can we give an explanation for *these* judgments? I think so.

If *Rational-Omniscience* is true, then if we ask what Bill should do *when he's not certain what credences are rational*, we're asking how someone who is already irrational should proceed. This can be a problematic question at times, but if *Rational-Omniscience* is true, we are talking about a *very* idealized notion of rationality. And if rationality is *so* idealized, it had better turn out that we can make *some* true judgments about what agents who aren't rationally omniscient should do and believe. So it seems, at very least, that when a non-ideal agent *knows* what choice is rationally required, we can correctly judge that it is rational for her to make that choice.[21] This is a very weak constraint on the correctness of rationality judgments about non-ideal agents.

What does this mean for Bill? Even if Bill isn't certain what credences are rational, so long as he knows that a perfectly rational agent would choose the rationality pill over an irrationality pill, it follows from the weak constraint above that *Bill* should choose the rationality pill as well. But *can* Bill know that this is what an ideally rational agent would choose? Yes. For as long as Bill knows that *Rational-Omniscience* is true, he can go through

---

[21] One might want to weaken this principle even further so that it says that if an agent knows that ideal rationality requires making a certain choice, *and* that a rational version of herself would *recommend* that she make that choice, then that is the choice that it is rational for her to make. This modification won't make any difference for what follows.

the reasoning above to convince himself that an ideally rational agent would choose the rationality pill (and the total evidence pill).

Still, you might worry that our intuitions about the game show cases weren't based on Bill's acceptance of any controversial principles like *Rational Omniscience*,[22] and so we shouldn't be relying on *Rational Omniscience* in explaining these judgments. In fact, however, I will argue that our judgments about such cases *should* depend on what we think about *Rational Omniscience*. As we'll see, commonly accepted views that reject *Rational Omniscience* have the consequence that Bill *shouldn't* always choose these pills. Our inclination to judge that it is rational to prefer, for accuracy purposes, the credences that our total evidence rationalizes is, in fact, based on an implicit commitment to the thought that a rational agent is always certain about what is rational.[23] This will be argued for in section 3.

Let's take stock. I started out asking three questions: Why should we prefer rational credences to irrational ones? Why should we prefer credences rationalized by our total evidence to credences rationalized by a proper subset of the evidence? What is the constitutive connection between rationality and accuracy? The picture I described in this section was based on taking, as our answer to the constitutive question, Best-Plan-to-Follow: the claim that the rational credences are those that accord with the plan that a rational agent would, a priori, regard as accuracy optimizing. I then showed that if we assume that Pria (the perfectly rational ignorant agent) is certain about which plan is the rational plan, there is an argument to be made from Best-Plan-to-Follow to RatPref and TotPref.

---

[22] Thanks to Christopher Peacocke for raising this worry.

[23] I know of no arguments for principles like RatPref and TotPref that don't ultimately rely on *Rational-Omniscience.* The only argument in the literature that I am familiar with for a principle like RatPref is the argument in Horowitz's, *op. cit.*'s discussion of "the value question." But her argument only works if *Rational Omniscience* is assumed. The only argument in the literature that I am familiar with for a principle like TotPref is an argument in Good, I.J. (1967) "On the Principle of Total Evidence." *British Journal for the Philosophy of Science* 17(3): 319-321. However, Good's argument only successfully establishes that you should assign greater expected accuracy to the credences based on your total evidence, than to credences based on a subset of that evidence, *before getting the evidence*. It doesn't justify the claim that, *once you get the evidence*, you should assign greater expected accuracy to the credence that's rational given the total evidence than to the credence that's rational given a proper subset of this evidence. For this, you need *Rational-Omniscience*. Good does suggest a reply to a similar worry, but it is unsatisfactory. Diagnosing the flaw in Good's response would take us beyond the scope of this paper, but it's also not necessary. For we'll see later that, without *Rational-Omniscience,* we can get straightforward counterexamples to TotPref.

3. Why One Might be Unhappy with Best-Plan-to-Follow: Higher Order Evidence

The aim of this section is to show why one might want to reject the picture that I described in the previous section. To begin, consider:

HYPOXIA[24]

While flying an airplane, you perform some calculations to determine whether or not you have enough fuel to make it to Hawaii. You suddenly get a message from ground control: "Dear pilot, you are flying at an altitude which makes you susceptible to hypoxia, a condition that impairs your ability to reason properly. The judgments[25] people at your altitude make concerning how far they can fly are correct only 50% of the time."

How confident should you be that you have enough fuel to make it to Hawaii? Plausibly, you should be only 50% confident. I will call the view that motivates this sort of thought "calibrationism." More precisely, let the antecedent probability that you assign to your judgment being correct in the circumstances in question your "expected degree of reliability."

Calibrationism: If your expected degree of reliability concerning whether p is 0.5, and you judge that p, you should be 0.5 confident in p.

Two notes: First, calibrationism is by no means uncontroversial, and my purpose here isn't to defend the view. All I am claiming is that the view has some intuitive plausibility. It has also been defended by many people in the debate about higher order evidence.[26] Since, as I

---

[24] This case is a variant of a case from Elga, A. (ms.) "Lucky to Be Rational."

[25] The agent's judgment is just the proposition that the agent is (or would be) most confident in on the basis of the first order evidence alone.

[26] Defenders include Roush, S. (2009) "Second Guessing: A Self-Help Manual." *Episteme* 6(3): 251-68; Feldman, R. (2009) "Evidentialism, Higher Order Evidence and Disagreement" *Episteme* 6(3):294-312; Christensen, D. (2010) "Higher Order Evidence" *Philosophy and Phenomenological Research* 81(1):185-215; Horowitz, S. And Sliwa, P. (forthcoming) "Respecting *All* the Evidence" *Philosophical Studies*; Elga, A. *op. cit.* and Vavova, K. (ms.) "Irrelevant Influences." Dissenters include Lasonen-Aarnio, M. (2010) "Unreasonable Knowledge" *Philosophical Perspectives* 24(1):1-21; Kelly, T. (2011). "Peer Disagreement and Higher Order Evidence" in A. Goldman & D. Whitcomb (eds.) *Social Epistemology: Essential Readings*. Oxford University Press; Titelbaum, *op. cit.* andWeatherson, B. (ms.) "Do Judgments Screen Evidence?". See also Schoenfield, M. (forthcoming) "A Dilemma for Calibrationism" *Philosophy and Phenomenological Research* and White, R. (2009) "On Treating Oneself and Others as a Thermometer." *Episteme* 6(3):233-250.

will argue, the view is inconsistent with Best-Plan-to-Follow, we may eventually have to make a choice about whether to give up calibrationism or give up Best-Plan-to-Follow. Both calibrationism and Best-Plan-to-Follow are prima facie plausible and have garnered much support in the literature so the decision about which to abandon should be made with care. Second, the version of calibrationism I presented above is simpler than some, and weaker than some, but the various bells and whistles that one might add need not concern us here.[27] I'll be applying calibrationism only to the most uncontroversial cases in which it is most plausible for such a principle to apply.

So why is the picture sketched in the previous section unavailable to the calibrationist? Let's begin with Best-Plan-to-Follow.

Why the Calibrationist Must Reject Best-Plan-to-Follow

To see why calibrationists have to reject the claim that the rational credences are the accuracy optimizing ones, recall that Pria is looking for an epistemic plan which optimizes the expected accuracy of the credences an agent following this plan would adopt. Now, put yourself in Pria's shoes. Suppose you're trying to decide, a priori, how to program your robot's responses to some ordinary meteorological evidence E. You conclude that the accuracy optimizing response would involve the robot believing that it will rain if given evidence E. You now consider the evidence that consists of E and also higher order evidence about the robot's cognitive functioning. If it's up to you to determine what credence function the robot adopts in these circumstances, you should want it to *still* believe RAIN. For you shouldn't think that it is any less likely to rain in circumstances in which the robot has E, in addition to the evidence about its cognitive functioning, than it is to rain in the circumstances in which the robot has E. Since the calibrationist thinks that what is *rational* given these two bodies of evidence differs, the calibrationist must reject Best-Plan-to-Follow.

Here is another way of making the point: Recall that the accuracy-optimizing plan is the plan, $\mathcal{P}$, that maximizes the expected accuracy of the credence functions *that $\mathcal{P}$ recommends* given a body of evidence. So, as far as the accuracy-optimizing plan is

---

[27] In particular, I am only assuming that one should calibrate concerning whether p when one's expected degree of reliability is 0.5, so as to remain neutral between the two versions of calibraitonism that I discuss in Schoenfield *ibid*.

concerned, whether an agent will likely *succeed* at adopting the credence function in question is completely irrelevant. Thus, if we're considering, a priori, the plan that issues the most expectedly accurate *recommendations*, one that disregards higher order evidence will be more expectedly accurate than one that accounts for it. The calibrationist will thus think that Best-Plan-to-Follow gives an oversimplified answer to the constitutive question. It attends only to how good a plan is *as a plan to follow*, and so disregards important evidence the agent might receive concerning the likelihood that she will actually *succeed* at following such plans.[28]

Since the calibrationist must reject Best-Plan-to-Follow, we might wonder how the calibrationist *should* answer the constitutive question. We will get to this in section 4. But first we need to consider how the calibrationist might respond to the preference questions: what can she say about RatPre*f* and TotPref?

The first thing to note is that calibrationists will have to reject *Rational Omniscience.* After all, plausibly it is *because* in certain cases, agents should become unsure about what their initial body of evidence supports, that the higher order evidence should lead to a revision of their beliefs. So the arguments I gave for RatPref and TotPref are not available to the calibrationist. But can she accept them anyway? Let's begin with TotPref.

Why The Calibrationist Must Reject TotPref

The calibrationist will have to reject TotPref. In other words, the calibrationist will think that sometimes we should prefer having credences supported by a proper subset of our evidence to the credences supported by our total evidence. To see why, suppose you find yourself in a scenario like HYPOXIA. Let's call the evidence that you get from ground control about your susceptibility to hypoxia your "higher order evidence" and the evidence you had before getting this evidence – the evidence you gained from looking at the plane's dials, gauges and maps – your "first order evidence." Suppose that, before you received the higher order evidence, you were confident that you had enough fuel, but upon hearing the message from ground control, you have reduced your confidence to 0.5. Still, you should regard the expected accuracy of the credences that satisfy the description "the credences

---

[28] I give a formal proof for the claim that calibrationism is inconsistent with Best-Plan-to-Follow in Schoenfield, M. (ms.) "An Accuracy Based Approach to Higher Order Evidence."

that are rational given my first order evidence" to be greater than the expected accuracy of the credences that satisfy the description "the credences that are rational given my total evidence." (Note that I'm *not* claiming that you should think that the credences *you would have* on the basis of the first order evidence alone are more expectedly accurate than the credences that are rational given your total evidence).

If this doesn't seem intuitive to you, think about what credences you'd prefer to have your *actions* guided by (and recall, that this is the sense of "prefer" that is relevant in this paper). Let's suppose for simplicity that your evidence either entails that you have enough fuel or that you don't (but some complicated reasoning is required to derive this entailment). If your actions are guided by the credences that, according to calibrationism, your *total* evidence supports, you'll have a 0.5 credence that you have enough fuel, and so you're going to play it safe and go home. This would be a very sad way to commence your much-anticipated vacation! On the other hand, if you could choose to have your actions guided by the credences that are *in fact* rational given only the first order evidence, you would only return home if, *in fact,* you didn't have enough fuel. (Recall that your evidence entails either that you have enough fuel or that you don't). If you did have enough, you would keep flying to Hawaii and be sipping a cocktail on the beach within a few hours. The expected value of either having a shot at arriving safely in Hawaii or returning home is greater than the expected value of a certain return home. So, in terms of the credences that you'd want your *actions* guided by, you should prefer the credences rationalized by a proper subset of your evidence to the credences rationalized by the total. This is inconsistent with TotPref.[29]

Calibrationism and RatPref

Whether or not the calibrationist will have to reject RatPref depends on what kind of calibrationist one is. On the view that I have called in other work[30] "E-calibrationism," rational agents should sometimes assign greater expected accuracy to a particular credence function g than to the credence function under the description "the rational credence given

---

[29] See also Horowitz, S. (ms.) "Predictably Misleading Evidence" for an argument for the claim that we should expect higher order evidence to be predictably and systematically misleading.

[30] Schoenfield, M. (forthcoming). "A Dilemma for Calibrationism." *op. cit.*

my evidence." But arguing for this in detail will take us beyond the scope of this paper.[31] All I will say about RatPref here is that calibrationism leaves the principle unmotivated. So far, we don't have an account of *what it is about rationality* that could explain why, for accuracy purposes, we should prefer rational credences to credences that we know are irrational.

4. Best-Plan-to-Make

The aim of this section is to describe an alternative connection between rationality and accuracy that is consistent with (and, as a bonus, motivates!) calibrationism. For reasons of space, I will not be able to do justice to the account's benefits *or* its problems. The aim here is just to put the proposal on the table and suggest that something like it lies at the heart of calibrationism.

This alternative account takes the idea of an "epistemic plan" quite literally. But this time, the planner is not asking *what's the best plan to follow?* Rather, she is asking, *what's the best plan to make*? In cases in which one is taking seriously the possibility that one might fail to conform to a plan that one makes these two questions will yield different answers.

Indeed, in ordinary life, when we compare plans, we are usually more interested in the expected results *of making a plan* than we are in the expected results of *conforming to it*

---

[31] Here's a quick version of the argument: Suppose one adopts the position that I have called "E-Calibrationism" in Schoenfeld, *ibid.* On this view, one's credence in p, *when p is the proposition that the first order evidence best supports* should equal one's expected degree of reliability. (E-calibrationism is motivated by the idea that one's credences shouldn't depend *only* on one's judgment and one's higher order evidence: one's first order evidence should play a role as well. See Schoenfeld *ibid.,* Horowitz and Sliwa *op. cit,* Weatherson *op. cit.* and White *op. cit.* for a discussion of why this is an important desideratum for a calibrationist theory). In that paper, I describe the following type of case: I judge that tomorrow will be sunny on the basis of meteorological evidence (this is my first order evidence). The oracle then comes down from the heavens and tells me the following: "Whenever you get strong meteorological evidence that it's going to rain tomorrow, you become so depressed that you end up inventing very silly arguments for the conclusion that it will be sunny. On the other hand, if you get strong meteorological evidence that it will be sunny, your fear of disappointment leads to your judging that it will be rainy!" In this case, my expected degree of reliability will be close to zero (for I believe that the judgments I make *on the basis of the first order evidence alone* are *anti*-reliable). So, according to E-calibrationism, the rational credence *in the proposition that the first order evidence best supports* will be close to zero. In this case, I'll think that the proposition that the first order evidence best supports is probably the proposition that it will rain, since I judged it will be sunny. So, if I accept E-calibrationism, I'll think that the rational credence function is one that assigns a very *low* credence to the proposition that it will rain. But I, of course, should prefer to adopt a credence function that assigns a *high* credence to the proposition that it will rain since rain is, after all, is what I think the meteorological evidence best supports! Thus if offered a choice between adopting a high credence in rain or the rational credence in rain, I'd prefer to adopt the high credence in rain, thereby violating RatPref.

(though frequently these two things go together).[32] For example, it might be that the vacation plan that is best to *conform* to is one in which I fly a spaceship to the moon. Nonetheless, that is surely not the best plan *to make*, even if it psychologically possible for me to make such a plan. To see why the plan to go to the moon is a bad plan to make, consider what I expect will occur if I make the plan to go to the moon. Perhaps I will spend my vacation filing various petitions and requests with NASA. Or maybe I'll try (and fail) to build a spaceship. Or perhaps I'll just realize that there is no way that I'll be able to fly a spaceship to the moon and, since I'll have made no other plans, I'll spend the week brooding at home. But whatever I think the expected consequences of making the moon plan are, plausibly, the expected value of making a plan to go to, say, Paris is higher.

Note that not all plans that are unlikely to succeed are bad plans. I may plan to quit smoking even if I think I'm unlikely to succeed. This might be because I think there's *some chance* I'll succeed and, even if I don't, I'll end up smoking less than I would if I didn't make this plan. You might think that we can even make plans that we'd *prefer* not to conform to. For example, if I am always running half an hour late, and I want to leave my house at 10:30, I will plan to leave at 10. This is not because I think it would be good to leave at 10. It's because I think that the result of *planning* to leave at 10 will be leaving at 10:30, and *that* is a desirable outcome.

One might take a similar stance when one considers the expected value of various epistemic plans – plans about how to form beliefs. You might be more interested in the expected accuracy of the beliefs you would form as a result *of having made a certain epistemic plan* than the expected accuracy of the beliefs that you would form by *conforming to a certain epistemic plan*. If so, you might think of the rational plan as the epistemic plan that is best *to make*, rather than the epistemic plan that is best to follow.

On the planning picture, here is how theorizing about rationality would proceed: one thinks (a priori) about some body of evidence E, (or some evidence of a certain type) and one asks one's self: "what should my belief plan be for circumstances in which I get E?" Depending on what E is, when you are considering this question, you may need to be open to the possibility that some of the plans you could *make* would have a low chance of being

---

[32] For a detailed argument for a similar claim, see Hedden, B. "Options and the Subjective Ought," *Philosophical Studies* 158(2): 343-360.

followed. In engaging in this deliberation, one considers the following question: If I make plan P, and I get evidence E, what are the likely doxastic states that I would adopt in that situation, and how accurate are those states likely to be?

To see how this works it will be helpful to have a concrete case in mind, so let's return to the case of the hypoxic pilot. Consider two epistemic plans about how to respond to higher order evidence:

Right Reasons: If you have first order evidence E, and then get higher order evidence which suggests that your expected degree of reliability concerning whether p is 0.5, adopt the credence in p that's rational given E. (That is, ignore the higher order evidence).

Calibrationism: If your expected degree of reliability concerning whether p is 0.5, and you judge that p, have a 0.5 credence in p.

Right Reasons is a great plan to conform to, but it's not a good plan to make. Since you know that you'll be impaired in the relevant class of cases, you can predict that one of two things will happen as a result of planning to do what *Right Reasons* recommends. One possibility is that you'll throw up your hands and think: "I have no idea what the principle recommends in this case! I'm hypoxic!" If that is the result of making the plan, maybe you'll suspend judgment, or maybe you'll pick a random credence out of a hat. More likely, perhaps, is that, as a result of making the *Right Reasons* plan, you will end up believing the propositions that you *think* the first order evidence supports, and these beliefs will, despite your best intentions, be wrong 50% of the time. (Practically, this means a 50% chance of death in the hypoxia case!). So the expected accuracy (and the expected value!) of the results of *planning* to follow calibrationism is greater than the expected accuracy of the results of *planning* to have the attitude that the first order evidence supports. That is, one should assign greater expected accuracy to the belief states that would result from planning to calibrate than to the belief states that would result from planning to follow

Right Reasons.[33] Thus, an alternative way of thinking about how rationality and accuracy are connected, which is consistent with and, in fact, motivates calibrationism is the following:

> **Best-Plan-to-Make**: The rational epistemic plan is the one that a rational agent would choose, a priori, if she were aiming to maximize the expected accuracy of the credences an agent would adopt as a result of *making* the plan. The rational credences given a body of evidence are those that the rational epistemic plan recommends given that body of evidence.

There is much to be said about Best-Plan-toMake and, as I mentioned at the outset, I will not, in this paper, be able to do justice to its merits or its problems, but below I list a few benefits of the account and some of the challenges it will face.

5. Benefits of Best-Plan-to-Make

Benefit #1: Motivating Calibrationism

As I mentioned, one benefit of Best-Plan-to-Make for those who are sympathetic with calibrationism is that it motivates calibrationism. Indeed, I think that any calibrationist who favors an accuracy-driven approach to epistemology *must* accept something like Best-Plan-to-Make. For on accounts according to which we evaluate the merits of a principle by examining (even a priori) the results of *following* it, rather than *planning* to follow it, *Right Reasons* will be preferred. This is so even though *Right Reasons* (unlike the principle "believe the truth") makes recommendations about what agents should believe *as a function of their evidence alone*.

A nice feature of this way of motivating calibrationism is that it doesn't rely on the idea that the principles of rationality need to be such that an agent is always able to follow

---

[33] Robert Steel (ms.) "Peer Disagreement, Anticipating Failure, and Avoiding It" adopts a similar strategy in defending conciliatory responses to disagreement. Maria Lasonen-Aarnio (2010) "Unreasonable Knowledge." *Philosophical Perspectives* 24(1):1-21 suggests that there may be a benefit to having the *dispositions* that accord with calibrationism. I give a formal argument for the claim that planning to calibrate is the procedure that maximizes expected accuracy in Schoenfield, M. (ms.) "An Accuracy Based Approach to Higher Order Evidence."

them. Indeed, it is compatible with Best-Plan-to-Make that the principles of rationality are extremely demanding and difficult. All that Best-Plan-to-Make requires is that they *not* be such, that there's some *alternative* plan, from which we can, a priori, expect greater success. So the problem with *Right Reasons*, on this picture, isn't that it's a principle that agents won't successfully follow. The problem is that there's a principle that is incompatible with *Right Reasons* that one can expect, a priori, to lead to greater epistemic success.

Benefit #2: Planning and Metaepistemology

Although one needn't adopt any particular metaepistemological theory to be interested in Best-Plan-to-Make, recent expressivist views of justification sit nicely with such a picture. Karl Schafer[34] for example, independently motivates the idea that, in claiming that it's rational to believe p give E, one is expressing a commitment to the *plan* to believe p if given evidence E. If this metaepistemological theory is on the right track, it would make sense to endorse epistemic norms that constitute plans that are good to *make*.

Benefit #3: Planning and Epistemic Guidance

The plan-making account sits nicely with the idea that the principles of rationality should be *guidance giving*. Indeed, planning can be thought of as giving guidance to your future self. And just as, when we plan, we frequently focus on the plan that's best to *make,* when we guide, we frequently focus on the guidance that's best to *give* (rather than the guidance that's best to follow). For example, I might know that the route to the train station that it would be best for someone to take involves lots of twists and turns, back alleys and shortcuts. It's a beautiful route, and if you follow it, you'll get to the train station quickly. But if I'm giving directions to a confused stranger, directing her to follow this route may be bad advice. This is because, we might imagine, the expected result of my advising her to take the scenic shortcut will be that she'll get lost, confused, and miss her train. I should give the stranger a more direct and less scenic route, not because that's the route it would be best for the stranger to *take*, but rather, because I can expect better results from

---

[34] Schafer, K. (2014) "Doxastic Planning and Epistemic Internalism" *Synthese* 191(12):2571-2591. I discuss Schafer's view and the role that planning may play in epistemology in my "Internalism without Luminosity" (forthcoming). *Philosophcal Issues*.

advising her to follow the longer uglier route, than from advising her to follow the shorter scenic route.

6. Challenges to Best-Plan-to-Make

Challenge #1: How to Know What to Plan?

It may be unclear how one could evaluate the results of *making* certain plans (rather than following them) a priori.[35] After all, what an agent will end up believing as a result of making various plans will depend on all sorts of facts about the agent that Pria doesn't have access to.

This is a large issue that I will just briefly touch on. I think that what the defender of Best-Plan-to-Make should say is that what Pria has to go on, primarily, when considering which principles to recommend, is whatever information about the agent *is in the body of evidence* she is considering. For example, Pria's principles will recommend that an agent with evidence about hypoxia suspend judgment because the fact *that the agent is at an altitude which makes her unreliable* is included in the body of evidence that Pria is considering. Additionally, it's not necessary that Pria be *certain* what credence function an agent who makes a certain plan will adopt, given some body of evidence. To calculate the expected accuracy of *making* a plan, all that's needed is a probability distribution over the possible credence functions that an agent who has that evidence would adopt as a result of having made the plan.

I suspect that calibrationists, at least, will be happy with this result. For calibrationists frequently point out that their view isn't about what agents with limited capacities should believe; their view is about what agents *with evidence* about having limited capacities should believe. Thus, the only thing that will be relevant, from the calibrationist's perspective, is what evidence *the agent* has about her own capacities. And this is precisely the sort of information that Pria will be able to take into account when considering which plans are best to make.

---

[35] Thanks to Tom Kelly for raising this question.

Challenge #2: Worries about Epistemic Consequentialism

One might worry that Best-Plan-to-Make is going to be subject to worries having to do with the consequentialist spirit of the proposal. For example, suppose an agent receives evidence E, which entails (obviously) that p, but she also has evidence that tells her that if she believes ~p, God will make sure that all of her other beliefs are true. Wouldn't the best plan to make be one in which she believes ~p? But surely it is *epistemically irrational* to believe ~p in such circumstances! One might respond by adding constraints that will forbid tradeoffs between one's attitudes towards different propositions, or tradeoffs between one's attitudes at different times. But recent papers by Selim Berker,[36] Michael Caie,[37] Jennifer Carr[38] and Hilary Greaves[39] pose challenges to epistemic consequentialist frameworks, which they think will apply even once such restrictions are in place. I will not attempt to respond to these challenges here. All I want to say is that, as the authors above point out, these challenges are challenges to the recent accuracy driven approach to epistemology *in general*. My aim here is not to defend "accuracy-first" epistemology from such challenges but to compare different ways of thinking about how such an approach might be developed.[40]

Challenge #3: Is that All Rationality Amounts to?

You might worry that if epistemic rationality is what Best-Plan-to-Make says it is, rationality will not be able to serve many of the purposes we may have wanted it to serve. This is because we don't *only* use the notion of rationality in the context of first personal deliberation and planning. Furthermore, many nice principles we might have wanted (for example, RatPref and TotPref, probabilism, conditionalization[41]) may not be vindicated by such an account.

---

[36] Berker, S. (2013). "Epistemic Teleology and the Separateness of Propositions," *Philosophical Review* 122: 337-393.

[37] Caie, M. (2013). "Rational Probabilistic Incoherence," *Philosophical Review* 122 (4):527-575.

[38] Carr, *op. cit.*

[39] Greaves, *op. cit.*

[40] For a response to the consequentialist challenges see, for example, Konek, and Levinstein *op. cit.*

[41] Calibrationism, which is motivated by Best-Plan-to-Make, will arguably be inconsistent with both probabilism and conditionalization since it will sometimes require agents to have low confidence in logical truths and in propositions that their evidence entails.

I want to end by suggesting that we may be able to have our cake and eat it too. We can imagine Pria, up in rationality heaven, creating *two* epistemic plans. She creates the plan that is best to *follow*, and she creates the plan that is best *to make*. Perhaps the word "rational" is ambiguous between these two plans, and which plan we're interested in, in some particular instance, depends on our purpose. If we're in the business of planning how to go about forming beliefs from a first personal perspective, we are going to be interested in the plan that's best to make. But it may be that, for other purposes, the relevant norms are those that make up the epistemic plan that one would regard, a priori, as best to follow. RatPref and TotPref, it turns out, are true when we're thinking about rationality as the plan that is best to *follow* but not the plan that is best to *make*. On this blended approach, there is no single answer to the constitutive question. There are two important ways in which an epistemic plan can be motivated by accuracy-based considerations, and two notions of rationality corresponding to our interest in each of these plans.

7. Conclusion

I began by posing three questions: why should we prefer rational credences to irrational ones, why should we prefer the credences rationalized by our total evidence to those rationalized by a proper subset of it, and what, if anything, can be said about the connection that rationality bears to accuracy? I argued that *prima facie* plausible answers these questions can be given if we accept the view that the rational credences are those recommended by the plan that one would regard, a priori, as the best plan to follow. I then showed that calibrationism is inconsistent with this picture and some of the claims it motivated. This left us with a puzzle: how is the calibrationist thinking about the connection between rationality and accuracy?

I suggested an account on which what's important, for rationality, isn't the expected accuracy of *following* a plan. Instead, what we're interested in is the expected accuracy of the results of *making a plan*. Although this view has some attractive features, it also faces some challenges and may fail to capture everything that we find valuable about epistemic rationality. So, rather than abandoning the original claim about the connection between rationality and accuracy, we might instead recognize that there are two interesting

epistemic notions that serve two distinct roles: there is the plan that one should regard a priori as best to *follow*, and the plan that one should regard a priori as best to *make*. I believe that the difference between these two kinds of plans, and their importance, lies at the heart of the debate about higher order evidence.