Can accuracy motivate modesty?

Accuracy-first epistemology aims to show that the norms of epistemic rationality can be derived from the effective pursuit of accuracy. This paper explores the prospects within accuracy-first epistemology for vindicating "modesty": the thesis that ideal rationality permits uncertainty about one's own rationality. I give prima facie arguments against accuracy-first epistemology's ability to accommodate three forms of modesty: uncertainty about what priors are rational, uncertain about whether one's update policy is rational, and uncertainty about what one's evidence is. I argue that the problem stems from the representation of epistemic decision problems. The appropriate representation of decision problems, and corresponding decision rules, for (diachronic) update policies should be a generalization of decision problems and decision rules used in the assessment of (synchronic) coherence. I then show that the appropriate generalization allows for rational modesty.

1 Background

1.1 Accuracy-first epistemology

According to accuracy-first epistemology, the norms of epistemic rationality are the norms of effective pursuit of accuracy. Accuracy-first epistemologists, as I use the term, endorse the following principles:

Alethic vindication. The ideal credence function at a world w is the omniscient credence function at that world: the credence function v_w such that for all relevant propositions P,

$$v_w(P) = \begin{cases} 1, & \text{if } P \text{ is true at } w \\ 0, & \text{otherwise} \end{cases}$$

Perfectionism. The epistemic utility of a credence function is represented by its closeness (by some appropriate measure) to the ideal credence function.

Epistemic decision theory. An agent is epistemically rational just in case her credences and their evolution conform to appropriate epistemic decision rules (e.g. maximize expected epistemic utility; avoid epistemic utility dominance).

Combining alethic vindication with perfectionism yields the result that the epistemic utility of a credence function is its gradational accuracy: its proximity to the truth, by some appropriate measure. Let $\mathcal{W}$ a set of worlds, F be a boolean algebra over $\mathcal{W}$, $\mathcal{C}_F$ be the set of credence functions over F, and $\mathcal{P}_F \subset \mathcal{C}_F$ be the set of probability functions over F. **Global accuracy measures** ($\mathfrak{a} : \mathcal{C}_F \times \mathcal{W} \to \mathbb{R}$) assess the inaccuracy of credence functions at worlds. **Local accuracy measures** ($\mathfrak{a}_l : \mathcal{C}_F \times F \times \mathcal{W} \to \mathbb{R}$) assess the inaccuracy of credences in individual propositions at worlds. There is controversy over the class of appropriate accuracy measures; they are typically held to have the following properties.

Truth-directedness. For credence functions c and c', if for all $p \in F$, either $c'(p) \geq c(p) \geq v_w(p)$ or $c'(p) \leq c(p) \leq v_w(p)$, and for some $p \in F$, $c'(p) > c(p) \geq v_w(p)$ or $c'(p) < c(p) \leq v_w(p)$, then $\mathfrak{a}(c, w) > \mathfrak{a}(c', w)$.

Separability. $\mathfrak{a}(c, w) = \sum_{p \in F} \mathfrak{a}_l(c, p, w)$.

Strict propriety. For every $c \in \mathcal{P}_F$ and every $c' \in \mathcal{C}_F$ s.t. $c' \neq c$, $\sum_{w \in \mathcal{W}} c(w)\mathfrak{a}(c, w) > \sum_{w \in \mathcal{W}} c(w)\mathfrak{a}(c', w)$.

I will assume that if accuracy-first epistemology is correct, then ideally rational agents are not ignorant of the correct epistemic decision rules or of which functions are accuracy measures. For example: if maximizing expected utility is necessary for rationality, then ideally rational agents accept that maximizing expected utility is necessary for rationality; if epistemic utility functions must be truth-directed, they know that epistemic utility functions must be truth-directed.[1] Rational uncertainty of rational decision rules is very, very hard to make sense of, as the literature on normative uncertainty demonstrates.[2]

I will also assume that rational agents choose epistemic options that maximize expected accuracy.

1.2 Modesty

Whether an agent is rational is a contingent fact that depends on the state of her hardware. For example, agents who are rational at t may have their hardware mal-

[1] This paper can be interpreted as defending a conditional conclusion: if ideally rational agents necessarily know which epistemic decision rules and utility functions are appropriate, then…

[2] See Sepielli (2014) for an illustration of the demands of characterizing how rational uncertainty about norms of practical rationality might be possible.

function at t', or may receive (misleading) evidence that their hardware is malfunctioning at t. Example:

> **Agnosticillin.** Jane currently has credence .5 in the hypothesis h, on the basis of total evidence e. Then she's told by a reliable friend that her tea was almost certainly drugged with agnosticillin. People drugged with agnosticillin will tend to have credences that are too high or too low given their evidence. Agnosticillin is in no way introspectively detectable. Agnosticillin does not hamper people's ability to detect their own credences and Jane knows what her credences are. Jane is lucky: she was not drugged, but she has no way of knowing this. Jane is, in fact, an ideally rational agent.
>
> Assessment: Jane should be uncertain about whether her credence in h is rational on her evidence.

Cases like this have been used to motivate the thesis Modesty:

Modesty. Ideally rational agents can be uncertain of their own rationality.

This thesis is neutral with respect to whether rational higher-order uncertainty should impact first-order credences. It is endorsed by both "level-bridgers" and "level-splitters." Level-bridgers (e.g., Christensen 2007, 2009, 2010a,b; Elga 2007, 2013; Horowitz 2014) believe that there are rational constraints on combinations of lower- and higher-order credences, so that higher-order uncertainty can impact what first-order credences are rationally permissible. Level-splitters (e.g., Williamson 2011, 2014; Weatherson manuscript; Lasonen-Aarnio 2010) accept the possibility of rational higher-order uncertainty but treat it as irrelevant to first-order uncertainty.[3]

Note that even if an ideal agent disposed to satisfy reflection-like principles that demand coherence between lower-order and higher-order credences, they may be stably modest. Suppose Jane is disposed to level-bridge in the face of higher-order evidence. Since she has no more reason to suppose her credence is too low than that it's too high, she has no reason to adjust her credence in h in response to her higher-order evidence. But she is also in no position to be confident that her response to her higher-order evidence is rational. After all, she reasons, suppose she should have had credence .7 in h. Then, upon receipt of her higher-order evidence, she should

[3] The thesis is not uncontroversial and is denied by, e.g., Titelbaum (manuscript).

not have adjusted her credences, and should have ended up with credence .7 in h, instead of her actual credence of .5. Similarly for any other credal assignment.

Modesty may be generated by uncertainty about the demands of rationality in general, or about the demands of rationality given one's evidence, or about what one's evidence is, or about one's own epistemic states. There are different varieties of uncertainty about the demands of rationality in the Bayesian tradition:

1. *Prior uncertainty:* uncertainty about which ur-priors are rational
2. *Update uncertainty:* uncertainty about what update policy is rational, given a body of evidence
3. *Evidence uncertainty*: uncertainty about what one's evidence is

Each of these forms of uncertainty is normative uncertainty.[4] A fourth form of uncertainty that may yield modesty is not a form of normative uncertainty, but is relevant to our discussion:

4. *Introspective uncertainty:* uncertainty about what one's credences are or how one updates

The focus of this paper is modesty that generated by normative (epistemological) uncertainty rather than introspective uncertainty. We therefore focus on the modified thesis:

Transparent Modesty. Ideally rational agents can be uncertain of their own rationality without being uncertain of their own doxastic attitudes.

1.3 Epistemic options

We assumed that a rational agent A must prefer credences that maximize expected utility by A's own lights. Given strict propriety, if A's credence function is probabilistic, then A will prefer her own current credence function over all other credence functions. So, one might ask, why should A ever update on new evidence? Wouldn't that involve moving to a credence function that A expects to be worse?

[4] It may not be obvious that evidence uncertainty counts as normative. I assume that it is. Evidence is a normative category: it's information that the agent is required to take into account.

It depends on what *epistemic options* are available to the agent.[5] Strict propriety requires probabilistic agents to prefer their own credence functions over all alternative options if all of the alternative options are credence functions. But what if there are other epistemic options?

Greaves & Wallace (2006) propose a different kind of epistemic option: what I'll call *credal gambles*. Credal gambles are functions from worlds to credence functions. Insofar as the agent doesn't know which world is actual, she may not know which credence function she will end up with if she takes a credal gamble.

Credal gambles can have higher expected utility by the lights of a probabilistic credence function than the option of maintaining that credence function.

> **Toy example.** Suppose there are exactly two possible worlds: w_1, where h is true, and w_2, where h is false. Suppose A is uncertain about h. She has the option of maintaining her current (probabilistic) credence function, which she knows is not maximally accurate. (After all, she is uncertain about h; if it were maximally accurate, she would be certain either about h or its negation—whichever was true.) And she has the option of taking a credal gamble, which will involve adopting credence 1 in h and 0 in $\overline{h}$ if w_1 is actual, and adopting credence 0 in h and 1 in $\overline{h}$ if w_2 is actual. Then the expected accuracy of the credal gamble is maximal. So it must have higher expected utility for A than the option of maintaining her current credences.

Therefore, strict propriety does not have the result that rational agents will never prefer to change their credences. Rational agents will prefer favorable credal gambles over maintaining their own credence functions.

One might worry: shouldn't rational agents always prefer—and take—the credal gamble that maps each world to the omniscient credence function at that world? Truth-directedness guarantees that this credence function maximizes accuracy. But it is uncontroversial that an agent is not irrational for failing to be omniscient. Greaves & Wallace argue that not all credal gambles are epistemic options. For example, the credal gamble that assigns the omniscient credence function of each world to each world is not (or not always) an epistemic option.

Greaves & Wallace restrict epistemic options (which they call "available acts") to a specific set of credal gambles. To do so, they localize update policies to specific

[5] So-called because the analogy to practical decision theory is illuminating; epistemic decision theorists do not presuppose epistemic voluntarism.

learning experiences. Suppose A expects at t to undergo some learning experience at some later t', but isn't sure what she'll learn. Let $\mathcal{E}$ be the set of propositions that she thinks might be her total evidence upon undergoing this learning experience. Greaves & Wallace stipulate that $\mathcal{E}$ must be a partition. Credal gamble U is an epistemic option for A at t just in case, for all $e \in \mathcal{E}$, for all $w, w' \in e$, $U(w) = U(w')$. In other worlds: U is an epistemic option for A at t just in case there's a function $U_{\mathcal{E}} : \mathcal{E} \to \mathcal{C}$ s.t. for all $w \in W$, if $w \in e$, then $U(w) = U(e)$. Intuitively: the credence function that U has you adopt is a function of your total evidence. It doesn't involve you being sensitive to information that you don't possess. The natural interpretation of epistemic options is that they represent the agent's plan for how to update her credences: if she learns e_1, she'll update to c_1. If she learns e_2, she'll update to c_2. And so on.

Greaves & Wallace (2006) prove that if a rational agent prefers to maximize expected accuracy relative to a strictly proper accuracy measure, then the agent will prefer optional credal gambles that update by conditionalization on the agent's total evidence.

2 The puzzle

2.1 Uncertainty about coherence?

The central question of this paper: Is rational transparent modesty compatible with accuracy-first epistemology?

Consider again the Agnosticillin case: Jane is uncertain about whether her credences are rational. So she is either uncertain about whether these credences are related to each other in a coherent way, or uncertain which credences are the rational response to her exogenous evidence.

Can a rational agent be transparently modest about her own coherence? Suppose Jane has a probabilistic credence function. Strict propriety ensures that her credence function maximizes expected epistemic utility relative to itself. So if Jane knows that her credence function is probabilistic, then she should immediately deduce that her credence function is coherent.[6] If Jane is only uncertain whether her credences are coherent, then she must only have introspective uncertainty. So she will not count as transparently modest.

[6] Here I assume that accuracy-first epistemology requires ideally rational agents' certainties to be closed under entailment. This follows from probabilism.

So if a rational agent is transparently modest, her uncertainty must not be uncertainty about the internal coherence of her attitudes. Instead, it must be uncertainty about whether her credences are the appropriate response to her evidence, or in the case of prior-selection, lack of evidence.

2.2 Does accuracy-first epistemology permit prior uncertainty?

An agent's ur-priors can be metaphorically characterized as the credences she assigns before receiving any evidence. If some probabilistic ur-priors are rationally impermissible, then accuracy-first epistemology says that their impermissibility is entailed by the fact that they violate some epistemic decision rule for the pursuit of accuracy. For example, Pettigrew (2016) argues that the correct decision rule for ur-prior selection is Maximin. Maximin requires rational agents to choose an option whose worst possible outcome is no worse than the worst possible outcome of any other option. Given a sigma algebra F of relevant propositions, there is exactly one probability function that satisfies Maximin with respect to accuracy: one that assigns equal credence to all of the strongest non-empty elements of F. In other words, Pettigrew argues, this credence function will satisfy a principle of indifference.

We assumed that rationality requires knowledge of the correct epistemic decision rules and knowledge of what constitutes accuracy. So again, the rational agent can immediately deduce which ur-priors are rationally permissible. So accuracy-first epistemology rules out rational prior uncertainty.

2.3 Does accuracy-first epistemology permit update uncertainty?

For the same reasons, accuracy-first epistemology rules out rational update uncertainty (uncertainty about what update policy is rational, given a body of evidence), except insofar as the relevant uncertainty boils down to either evidence uncertainty or introspective uncertainty.

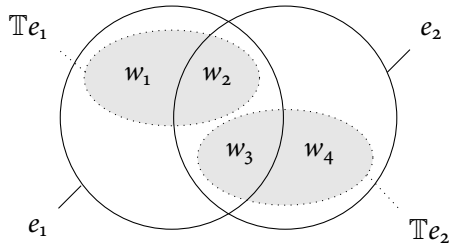
2.4 Does accuracy-first epistemology permit evidence uncertainty?

Schoenfield (2017) argues that accuracy-first epistemology rules out rational evidence uncertainty. The argument runs as follows:

Recall Greaves & Wallace's condition on epistemic options: A's learning experience is represented by the set of propositions $\mathcal{E}$ that A might learn at t. A credal gamble is an epistemic option just in case, for all $e \in \mathcal{E}$, U assigns the same credence

function to all worlds in e. What credence function the rational agent ends up with at a world will be a function of what her total evidence is at that world.

Greaves & Wallace presuppose that $\mathcal{E}$ is a partition. But let's suppose that in some circumstances an agent can regard it as possible that she learn e and possible that she learn $e' \neq e$ where e and e' are compatible. In such cases, if the agent's total evidence is e, her total evidence does not entail the proposition that e is her total evidence. Call this latter proposition $\mathbb{T}e$.



If e and e' are compatible but not identical, can they warrant different updates? Intuitively, yes. But this is impossible given the definition of an epistemic option. For reductio, suppose $U_{\mathcal{E}}(e) \neq U_{\mathcal{E}}(e')$. Select an arbitrary $w \in e \cap e'$. By the definition of epistemic options, $U(w) = U_{\mathcal{E}}(e)$ and $U(w) = U_{\mathcal{E}}(e')$. Contradiction.

So if the agent's predicted evidence is nonpartitional, then U cannot be a function of her evidence propositions. It must be a function of a partition over W.

Because evidentialism is true[7], the correct theory of epistemic rationality will make an agent's rational update policy a function of the total evidence she receives. What credence should she adopt in $w \in e \cap e'$? It depends on what total evidence she in fact receives from her learning experience in w. In all worlds where an agent's total evidence is e, epistemic options should assign the same credence function. But that means that the optional credal gambles will be those that are functions, not of $\mathcal{E}$, but of $\mathbb{T}\mathcal{E} = \{\mathbb{T}e : e \in \mathcal{E}\}$. Even if $\mathcal{E}$ is nonpartitional, $\mathbb{T}\mathcal{E}$ is partitional.

But note: for any partition Π s.t. all epistemic options assign uniform credence functions within the cells of Π, the epistemic option that maximizes expected accuracy will be one that is omniscient about the propositions in Π.[8]

Schoenfield concludes, accuracy-first epistemology requires that A prefer that if her total evidence is e, she be certain of $\mathbb{T}e$. Schoenfield shows that given Greaves

[7] Though not defended here.

[8] If the partition is maximally fine-grained, then all credal gambles are epistemic options, and then only the policy of updating to omniscience maximizes expected utility.

& Wallace's characterization of credal gambles as epistemic options, the update policy that maximizes expected accuracy is not, *pace* Greaves & Wallace, conditionalization: where c^* is the agent's prior credence function, and where $c^*(x \mid y) =_{df} \frac{c^*(x \wedge y)}{c^*(y)}$,

Conditionalization. Given total evidence e, adopt $c_e = c^*(\cdot \mid e)$.

Instead, Schoenfield shows, maximizing expected accuracy requires conditionalizing, not on e, but on $\mathbb{T}e$—even if e doesn't entail $\mathbb{T}e$. Call this rule "Schoenfield conditionalization":

Schoenfield conditionalization. Given total evidence e, adopt $c_e = c^*(\cdot \mid \mathbb{T}e)$.

So, given Greaves & Wallace's characterization of epistemic options, accuracy-first epistemology rules out rational evidence uncertainty.

Objection: Greaves & Wallace's result, and Schoenfield's generalization, only require that rational agents synchronically prefer, in advance, to update in certain ways on their total evidence. For all that, the agent may update in a different way, and end up with a probabilistic credence function that both maximizes expected utility by its own lights and exhibits evidence uncertainty.

Reply: insofar as accuracy-first epistemology is capable of supporting evidentialism at all, it will have to require that rational agents not only prefer some credal gambles over others as their update policies, but in fact conform to the update policies that they prefer. Update policies are the only moving part in the apparatus that is sensitive to evidence at all. Let c_e be the credence function that is supported by A's total evidence. Given A's epistemic situation, the only way to ensure that c_e maximizes expected accuracy by A's lights is to ensure that A has already updated to c_e (given strict propriety). How, then, is A's credence function specifically constrained by e? The constraint should come not from coherence but from update: conforming to a diachronic update policy that, by A's lights prior to receiving e, maximized expected accuracy.

2.5 Whither transparent modesty?

Problem: Accuracy-first epistemology seems to rule out all forms of transparent modesty. But examples like the Agnosticillin case suggested that transparent modesty is possible.

3 Lessons from introspective uncertainty

3.1 Does accuracy-first epistemology permit introspective uncertainty? First pass: no.

Does accuracy-first epistemology permit introspective uncertainty? At first pass: no.

Suppose some epistemic option U^* will sometimes generate a credence function with introspective uncertainty. That is, for some proposition p, it'll assign credence n, but will assign credence less than 1 to the proposition (*):

(*) My credence in p is n.

If the agent adopts this credence function in all worlds compatible with her evidence, then (*) is true at every world compatible with her evidence. But if a proposition is true at every world compatible with her evidence, then given truth-directedness, any credence function that maximizes expected accuracy will assign it credence 1.

So accuracy-first epistemology seems to rule out the possibility of introspective uncertainty.

3.2 Does accuracy-first epistemology permit introspective uncertainty? Second pass: yes.

The argument for why introspective uncertainty would be prohibited depended on what Carr (2017) called a "consequentialist" version of epistemic decision theory. Consequentialist epistemic decision theory functions identically to practical decision theory, except that it imposes restrictions on the space of options (epistemic options) and utility functions (accuracy measures). The logical space of its decision problems is the space of possible worlds. We can represent its decision problems, as usual, using partitions of options and partitions of possible states of the world. The possible states, orthogonal to the agent's acts, that the agent is uncertain about must be coarse enough to be compatible with multiple epistemic options. Decision problems can be represented with decision matrices, where columns represent possible states of the world and rows represent possible acts. A simple example:

	s_1	s_2
c_1	w_1	w_2
c_2	w_3	w_4

Here, w_1 and w_2 are worlds in which the agent adopts c_1.

Contrast consequentialist epistemic decision theory with nonconsequentialist epistemic decision theory—a form of decision theory tacitly employed by many accuracy-first epistemologists and necessary for results like Joyce's (1998; 2009) accuracy-dominance argument for probabilism, Greaves & Wallace's (2006) expected utility argument for conditionalization, Pettigrew's (2012; 2013) various arguments for the principal principle, and so on. These results do not hold up in consequentialist epistemic decision theory, and probabilism, conditionalization, and the principal principle are all subject to rational violations (Greaves 2013; Caie 2013; Carr 2017).

Nonconsequentialist epistemic decision theory is nonconsequentialist in the sense that it does not assess credence functions in terms of the epistemic utility gained as a consequence of the agent's adoption of those credence functions. Each option is assessed at all worlds—including worlds in which that option is not selected. This requires using a finer grained logical space, which allows for a different representation of epistemic options and epistemic decision problems.

	w_1	w_2	w_3	w_4
c_1	$\langle c_1, w_1 \rangle$	$\langle c_1, w_2 \rangle$	$\langle c_1, w_3 \rangle$	$\langle c_1, w_4 \rangle$
c_2	$\langle c_2, w_1 \rangle$	$\langle c_2, w_2 \rangle$	$\langle c_2, w_3 \rangle$	$\langle c_2, w_4 \rangle$

The logical space needed for this basic form of nonconsequentialist epistemic decision theory is a set of world–credence function pairs. Hence c_1 can be assessed as more or less accurate that c_2 at w_4—a world in which the agent in fact adopts c_2.

Distinguish between a *credence function* (a mathematical object; notation: c) vs. *an agent's act of possessing or adopting a credence function* at a time (a proposition: the set of worlds in which the relevant agent adopts the credence function at the relevant time; notation: $\mathbb{A}c$). Each is assessable for accuracy: c's accuracy *at* a world

vs. $\mathbb{A}c$'s accuracy *in* a world.

We can define an accuracy measure for $\mathbb{A}c$ as follows: for a set of worlds $s \subseteq \mathbb{A}c$, $\mathfrak{a}^*(\mathbb{A}c, s) = \mathfrak{a}(c, w)$ for all $w \in s$. If $s \not\subseteq \mathbb{A}c$, or if $\mathfrak{a}(c, w)$ isn't uniform across s, then $\mathfrak{a}^*(\mathbb{A}c, s)$ is undefined. Note that while c has a defined inaccuracy score at every world, $\mathbb{A}c$ does not. $\mathfrak{a}^*(\mathbb{A}c, \{w\})$ is defined only if w is a world in which the agent adopts c.

Nonconsequentialist epistemic decision theory diverges from traditional practical decision theories by redrawing the logical space of the decision problem.[9] Practical decision problems are organized in terms of self-locating information: an agent's A's options at a time t carve up the logical space in terms of the possible consequences of A's selecting each option at t. Nonconsequentialist epistemic decision theory in effect removes any self-locating information from the space of possible worlds. With no information about who A is in each world, the options before her cannot be represented by partitioning logical space according to what option she selects in each world.

Instead, the options partition logical space in a way that is orthogonal to which options the agent selects in each world. Indeed, to remove all self-locating information from the decision problem, the decision problem doesn't even assume that the agent exists in all of the worlds relevant to the decision. A's epistemic options do not partition W; they partition the enriched logical space $\mathcal{W} \times \mathcal{C}$.

For ease of exposition, it's helpful to distinguish the agent of the decision problem from her "counterparts" in each world.[10] A can assess the value (accuracy) of an option (c) at a world in which A's counterpart's credence function is $c' \neq c$. Importantly for present purposes: What credence A's counterpart has within w is no more relevant to which credence function A can assess at w than what credence function anyone else has within w.

So: c may assign n to p. But within some world–credence function pairs compatible with the c option are worlds in which A's counterparts don't assign credence n to p—that is, worlds in which ($\star$) is false. And so it won't necessarily maximize expected accuracy to be certain of ($\star$). Within nonconsequentialist epistemic decision theory, then, introspective uncertainty is not rationally prohibited.

---

[9] Briggs (2009) defends a form of practical analogue of nonconsequentialist epistemic decision theory and argues that choosing options that are dominated within this form of decision problem reveals incoherence, while mere Dutchbookability does not.

[10] This paper is neutral about the existence or nature of trans-world identity.

4 Epistemic decision theory for update

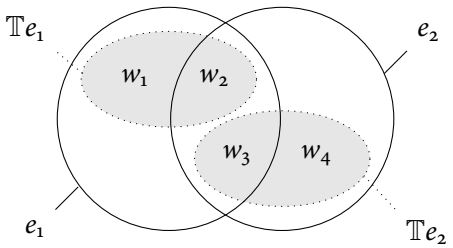
4.1 Generalization of nonconsequentialist decision theory

Nonconsequentialist epistemic decision theory allows for the construction of decision problems without self-locating information. The agent's credal options are represented as orthogonal to her counterparts' possible credences. So A's counterparts' credences play no more distinguished role in the decision problem than any other person's credences.

How might this representation of decision problems generalize from credence functions (constant credal gambles) to credal gambles in general? A's epistemic options should be functions of the total evidence A might receive. But because self-locating information is removed from the decision problems, A's epistemic options need not be functions of the total evidence that A's counterparts might receive. Crucially, in order for her decision problem to exclude self-locating information, it must be that whatever total evidence A's counterparts receive plays no more distinguished role in the decision problem than any other person's total evidence. How can this be possible?

We have to formally distinguish A's evidence from evidence her counterpart receives in any possible world. (Just as we distinguished A's optional credences from the credences A's counterparts take within any possible world.) But for update planning, A may still be uncertain about what evidence she will receive. So the state space for her decision problem cannot merely be $\mathcal{W}$: it must be $\mathcal{W} \times \mathcal{E}$. So the logical space for nonconsequentialist decision problems should be generalized to $\mathcal{W} \times \mathcal{E} \times \mathcal{C}$.

Return to the toy example of nonpartitional evidence, defined over possible worlds:



A decision problem for example might be represented as follows:

	$\langle w_1, e_1 \rangle$	$\langle w_2, e_1 \rangle$	$\langle w_3, e_1 \rangle$	$\langle w_2, e_2 \rangle$	$\langle w_3, e_2 \rangle$	$\langle w_4, e_2 \rangle$
U_1	$\langle w_1, e_1, c_1 \rangle$	$\langle w_2, e_1, c_1 \rangle$	$\langle w_3, e_1, c_1 \rangle$	$\langle w_2, e_2, c_2 \rangle$	$\langle w_3, e_2, c_2 \rangle$	$\langle w_4, e_2, c_2 \rangle$
U_2	$\langle w_1, e_1, c_3 \rangle$	$\langle w_2, e_1, c_3 \rangle$	$\langle w_3, e_1, c_3 \rangle$	$\langle w_2, e_2, c_4 \rangle$	$\langle w_3, e_2, c_4 \rangle$	$\langle w_4, e_2, c_4 \rangle$

Credal gambles, in this space, are functions from $\langle w, e \rangle$ pairs to credence functions.

4.2 Choice points

Greaves & Wallace's framework for assessing the expected accuracy of credal gambles rules out evidence uncertainty because of their representation of epistemic decision problems. I've argued that the appropriate representation of decision problems, and corresponding decision rules, for update policies should be a generalization of decision problems and decision rules used in the assessment of (synchronic) coherence.

Generalizing our representation of decision problems leaves open questions about how to generalize the corresponding representation of epistemic options, accuracy, and decision rules. Indeed, we're even left with choice points about the logical space for the decision problems. Finally, the epistemic decision problems, with their distinct representation of evidence that A receives from evidence that A's counterparts receive, stand in need of a philosophical interpretation.

My aim in this paper is to motivate a new framework for understanding epistemic decision problems for update policies, and to show that it accommodates rational transparent modesty. I will not take a stand on how these different choice points are best resolved. Below I'll explore some options, and then finally show how different options will yield the result that transparently modest update policies may be rationally permissible.

4.2.1 Epistemic options

The most straightforward generalization of Greaves & Wallace's epistemic options, tailored for nonconsequentialist epistemic decision problems, will make epistemic options the set of credal gambles that assign uniform credence functions to all $\langle w, e \rangle$ pairs that share an e coordinate.

Let $\mathbb{L}e_i$ be the proposition that includes all $\langle w, e \rangle$ pairs that have e_i as their e coordinate. Let $\mathbb{L}\mathcal{E}$ be the set of all propositions $\mathbb{L}e_i$ s.t. $e_i \in \mathcal{E}$, where $\mathcal{E}$, as usual,

represents the set of total evidence (possible worlds) propositions that A may learn as the result of her learning experience. $\mathbb{L}\mathcal{E}$ forms a partition even when $\mathcal{E}$ does not. (This is by design; as we saw, the assumption that possible learned evidence would need to be partitional entailed that rational agents exhibit no evidence uncertainty.) Then epistemic options may be represented as functions from $\mathbb{L}\mathcal{E}$ to credence functions. Alternatively, epistemic options may be more restricted.

We also face the question of whether the credence functions assigned by epistemic options are credence functions over possible worlds propositions or $\langle w, e\rangle$-propositions.

4.2.2 Accuracy measures

In the form of nonconsequentialist epistemic decision theory that is often presupposed in accuracy-first epistemology, the logical space for decision problems is not the space of possible worlds, but a space of world–credence function pairs. But the accuracy of a credence function c is not measured according to its proximity to the indicator function of a $\langle w, c\rangle$-pair. Instead, c's accuracy is measured according to its proximity to the indicator function of w. Here, there is no reason for c range over $\langle w, c\rangle$ propositions; instead, it can simply range over possible worlds propositions.

Our generalization for update must be tweaked: A is uncertain of what evidence she might receive (before receiving it; here I don't presuppose evidence uncertainty), and the evidence that she might receive is represented as orthogonal to the evidence that her counterparts receive across possible worlds. So A's credence function must range over $\langle w, e\rangle$-propositions. I discuss interpretations of this uncertainty below.

Here, we face a choice point over whether accuracy is to be measured according to proximity to possible worlds or $\langle w, e\rangle$-pairs. If we choose the former, simpler option, and if epistemic options assign credence functions that are defined over $\langle w, e\rangle$-propositions, then we must give up strict propriety.

Let p be a $\langle w, e\rangle$-proposition. Define $\mathbb{S}p$ as the set of $\langle w, e\rangle$-pairs such that if any pair in p has w_i as its world coordinate, then every pair with w_i as its world coordinate is in $\mathbb{S}p$. These are the analogues of possible worlds propositions in the new decision space: they do not make finer-grained distinctions. Any two credence functions that assign all the same probabilities to the $\mathbb{S}$-propositions that they are assigned over will have the same accuracy as each other at every world. Therefore neither will assign the other greater or lesser expected accuracy than itself. We can at best impose the weaker propriety constraint, adapted to the new space:

Propriety. For every $c \in \mathcal{P}_F$ and every $c' \in \mathcal{C}_F$ s.t. $c' \neq c$, $\sum_{\langle w,e\rangle \in \mathcal{W}\times\mathcal{E}} c(w,e)\mathfrak{a}(c,w) \geq \sum_{\langle w,e\rangle \in \mathcal{W}\times\mathcal{E}} c(w,e)\mathfrak{a}(c',w)$.

4.2.3 Logical space

Evidence is factive. For this reason, we might not treat all $\langle w, e\rangle$ pairs as possible, but instead rule out all $\langle w, e\rangle$-pairs s.t. $w \notin e$. Alternatively, we might allow such points in our logical space. Should these points be doxastic possibilities for agents? If not, we may separately derive a rational prohibition on assigning positive credence to any $\langle w, e\rangle$-pair where $w \notin e$.

4.2.4 Epistemic decision rules

Different variants of dominance avoidance and expected accuracy maximization may be appropriate depending on how the parameters for epistemic options, accuracy measures, and logical space are set.

4.2.5 Philosophical interpretation

Nonconsequentialist decision theories face a general challenge for how they should be philosophically interpreted.[11] Our extension faces these interpretive problems and others. In particular, we need a philosophical interpretation of what attitudes the believer takes toward $\langle w, e\rangle$-propositions. One suggestion: the agent's interaction with her evidence comes in two forms:

1. *Causal-normative role:* What evidence she actually receives should determine how she updates.

2. *Belief object role:* What evidence she might receive is a fact about the world that she can think about (in the same way that she can think about what evidence anyone else might receive).

For the purposes of update, these two roles may come apart. The first role is essentially self-locating; the second is not. The e coordinate satisfies the first role; facts about her counterparts' evidence in each w satisfy the second role.

The most conservative use of our framework will make the e component relevant only in the context of epistemic decision-making. Otherwise it will be invisible. In this case, it should have limited impact on the agent's credences in possible worlds

[11] See Carr (2017) for discussion.

propositions. The extent of the impact will be affected by choices of epistemic options, accuracy measures, logical space, and epistemic decision rules.

5 Proof of concept

Again, my aim is to introduce a representation of nonconsequentialist decision problems appropriate for update. I will not argue for any particular selections for the above choice points or show that given these selections, some update strategy or other is rational.

I also aim to show that this representation of epistemic utility, unlike Greaves & Wallace's, is capable of accommodating rational evidence uncertainty and hence transparent modesty. Below, I consider a few examples:

5.1 Example 1

First: suppose epistemic options are all and only functions from $\mathbb{L}\mathcal{E}$ propositions to credence functions. Suppose further that accuracy is measured relative to worlds, and that the logical space does not contain $\langle w, e\rangle$-pairs where $w \notin e$. Rational agents maximize expected accuracy, where the expected accuracy of an epistemic option U relative to a prior c^* is represented as follows:

$$\sum_{\langle w,e\rangle \in \mathcal{W}\times\mathcal{E}} c^*(w,e)\mathfrak{a}(U(w,e),w)$$

Now, one of the most compelling examples of rational evidence uncertainty is Williamson's (2011; 2014) case of the unmarked clock. Her is a simplified version:

> **Unmarked clock.** Jane is about to look at an "irritatingly austere" where the minutes and hours are entirely unmarked. The clock's minute hand moves in discrete one-minute steps. Jane knows that she will not be able to discern which exact minute the clock is pointing to: her visual evidence will not be fine-grained enough. Instead, she knows, what visual evidence she receives will leave a margin of error: if the clock in fact reads 4:21, she will only receive the evidence that the clock reads either 4:20, 4:21, or 4:22. In general, iff the time reads n, her evidence will be that the clock's reading is in $n \pm 1$ minute. Before seeing the clock, Jane sees every possible setting of the clock as equiprobable.

Suppose there are 1440 worlds: one for each reading of the clock. Let w_i be the world in which the clock reads i. At each w_i, Jane has evidence $e_i = \{w_{i-1}, w_i, w_{i+1}\}$.

How should Jane respond to whatever evidence she receives? Many[12] accept that, if Jane's evidence is e_i, Jane should conditionalize on e_i, becoming certain of it, but uncertain of which world in e_i is actual. Because it will be an open possibility, after learning e_i, that w_{i-1} is the actual world, it will be an open possibility for her that her evidence is not e_i but e_{i-1}. Indeed, if she conditionalizes on her prior, she will give each world in e_i equal probability, and so will be $\frac{2}{3}$ confident that e_i is not her evidence. Hence she will exhibit evidence uncertainty and, assuming she introspects her credences, will be transparently modest: uncertain of what her evidence is, and therefore whether her credences are rational on her evidence.

Jane's evidence is nonpartitional. Given Greaves & Wallace's representation of decision problems, the update strategy that is rational for Jane is Schoenfield conditionalization. This will require her to be certain not just of her evidence, but of the specific reading of the clock. This follows from the fact that for every e_i, $\mathbb{T}e_i \equiv \{w_i\}$.

Can our framework do better? Given the assumptions above, the Greaves & Wallace result entails that, within our new epistemic decision problems, any epistemic option that maximizes expected utility within this framework will be one that assigns credences over $\mathbb{S}$-propositions that are updated by conditionalization on $\mathbb{L}e$. (Other propositions do not impact accuracy.) Suppose that Jane's prior (before seeing the clock) distributes credence equally among the possible $\langle w_i, e_j\rangle$-pairs. Then the epistemic option that maximizes expected accuracy will assign equal probability to $\langle w_{i-1}, e_i\rangle$, $\langle w_i, e_i\rangle$, and $\langle w_{i+1}, e_i\rangle$. Since all three are worlds where e_i is true, and w_{i-1} and w_{i+1} are both worlds where $\mathbb{T}e_i$ is false, this epistemic option will be certain of $\mathbb{S}e_i$ but only have credence $\frac{1}{3}$ that $\mathbb{T}e$, as desired.

5.2 Example 2

The epistemic option that will maximize expected utility within this framework, given the assumptions in the previous subjection, will update by conditionalization on $\mathbb{L}e$. This will not always coincide with conditionalization on e among possible worlds propositions. When a rational agent receives evidence e_1, she'll update on $\mathbb{L}e_1$—a strictly stronger proposition than e_1. Her resulting credences may therefore violate conditionalization with respect to possible worlds propositions.

It's not obvious that this is a bad result. Gallow (2014, unpublished) and Bronf-

[12] Williamson (2011, 2014); Christensen (2010b); Elga (2013).

man (2014) have argued that in cases where an agent's future evidence is expected to be nonpartitional, or cases where the agent does not know what her evidence is, updating by conditionalization is sometimes irrational. Consider the toy example from section 2.4.

Suppose that the agent's prior c^* is divided evenly over $w_1, \ldots, w_4$. Then since w_2 and w_3 are both compatible with two evidence propositions, in our finer logical space, they'll each have to divide into two $\langle w, e \rangle$ pairs; we'll again split the agent's credence evenly.

	$\langle w_1, e_1 \rangle$	$\langle w_2, e_1 \rangle$	$\langle w_3, e_1 \rangle$	$\langle w_2, e_2 \rangle$	$\langle w_3, e_2 \rangle$	$\langle w_4, e_2 \rangle$
c^*	$1/4$	$1/8$	$1/8$	$1/8$	$1/8$	$1/4$

The left three boxes correspond the $\mathbb{L}e_1$ and the right to $\mathbb{L}e_2$. If the agent conditionalizes on $\mathbb{L}e_1$, her credences will update to $c^*(\cdot \mid \mathbb{L}e_1)$, which differs in possible worlds propositions from updating by conditionalization on e_1:

	w_1	w_2	w_3
$c^*(\cdot \mid \mathbb{L}e_1)$	$1/2$	$1/4$	$1/4$
$c^*(\cdot \mid e_1)$	$1/3$	$1/3$	$1/3$

Note that while this update violates conditionalization, it conforms to the alternative to conditionalization, ExCondi, defended in Gallow (unpublished). The extent of this consonance depends on the assignment of priors over the enriched logical space.

5.3 General considerations

If conditionalization on $\mathbb{L}e$-propositions doesn't conform to conditionalization over possible-worlds propositions, is it utterly unconstrained? No: at the level of possible worlds propositions, the resulting credence functions will conform to Jeffrey Conditionalization relative to some input partition. This partition will be non-trivial in any case where there are questions that $\mathbb{L}e$ is irrelevant to. Further constraints may come from motivated restrictions on the distribution of priors over relevant $\langle w, e \rangle$-propositions.

Other assumptions about the space of epistemic options, accuracy measures,

logical space, epistemic decision rules, and philosophical interpretation may be warranted. I have focused on these because they are the most conservative, not the most plausible, extensions of the traditional accuracy-first framework. There are possible restrictions on epistemic options that yield the result that conformity to conditionalization over possible worlds propositions maximizes expected utility. There are other possible restrictions that instead require Schoenfield conditionalization and prohibit evidence uncertainty. This representation of epistemic decision problems merely allows, rather than mandates, transparent modesty.

6 Conclusion

Various forms of higher-order uncertainty seem impossible within accuracy-first epistemology. Evidence uncertainty is ruled out by Schoenfield's result if epistemic decision problems are understood on the model presented in Greaves & Wallace. But this representation of decision problems appropriate for update doesn't make sense, given the understanding of epistemic decision theory needed to secure any of the classic accuracy-first results. There's a more general model for epistemic decision problems, and corresponding epistemic options, that does not require updating by conditionalization to evidence certainty. With this generalization, accuracy-first epistemology is able to accommodate, and even vindicate, transparent modesty.

References

Briggs, R.A. (2009). "Distorted Reflection." *Philosophical Review*, 118(1): pp. 59–85.

Bronfman, Aaron (2014). "Conditionalization and Not Knowing That One Knows." *Erkenntnis*, 79(4): pp. 871–892.

Caie, Michael (2013). "Rational Probabilistic Incoherence." *Philosophical Review*, 122(4): pp. 527–575.

Carr, Jennifer Rose (2017). "Epistemic Utility Theory and the Aim of Belief." *Philosophy and Phenomenological Research*, 95(3): pp. 511–534.

Christensen, David (2007). "Epistemology of Disagreement: The Good News." *Philosophical Review*, 116(2): pp. 187–217.

Christensen, David (2009). "Disagreement as Evidence: The Epistemology of Controversy." *Philosophy Compass*, 4(5): pp. 756–767.

Christensen, David (2010a). "Rational Reflection." *Philosophical Perspectives*, 24(1): pp. 121–140.

Christensen, David (2010b). "Rational Reflection." *Philosophical Perspectives*, 24(1): pp. 121–140.

Elga, Adam (2007). "Reflection and Disagreement." *Noûs*, 41(3): pp. 478–502.

Elga, Adam (2013). "The Puzzle of the Unmarked Clock and the New Rational Reflection Principle." *Philosophical Studies*, 164(1): pp. 127–139.

Gallow, J. Dmitri (2014). "How to Learn From Theory-Dependent Evidence; or Commutativity and Holism: A Solution for Conditionalizers." *British Journal for the Philosophy of Science*, 65(3): pp. 493–519.

Gallow, J. Dmitri (unpublished). "Updating for Externalists."

Greaves, Hilary (2013). "Epistemic Decision Theory." *Mind*, 122(488): pp. 915–952.

Greaves, Hilary, and David Wallace (2006). "Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility." *Mind*, 115(459): pp. 607–632.

Horowitz, Sophie (2014). "Epistemic Akrasia." *Noûs*, 48(4): pp. 718–744.

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

Joyce, James M. (2009). "Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief." In F. Huber, and C. Schmidt-Petri (eds.) *Degrees of Belief*, Synthèse, vol. 342, pp. 263–297.

Lasonen-Aarnio, Maria (2010). "Unreasonable Knowledge." *Philosophical Perspectives*, 24(1): pp. 1–21.

Pettigrew, Richard (2012). "Accuracy, Chance, and the Principal Principle." *Philosophical Review*, 121(2): pp. 241–275.

Pettigrew, Richard (2013). "A New Epistemic Utility Argument for the Principal Principle." *Episteme*, 10(1): pp. 19–35.

Pettigrew, Richard (2016). "Accuracy, Risk, and the Principle of Indifference." *Philosophy and Phenomenological Research*, 92(1): pp. 35–59.

Schoenfield, Miriam (2017). "Conditionalization Does Not Maximize Expected Accuracy." *Mind*, 126(504): pp. 1155–1187.

Sepielli, Andrew (2014). "What to Do When You Don't Know What to Do When You Don't Know What to Do…." *Noûs*, 48(3): pp. 521–544.

Titelbaum, Michael G. (manuscript). "In Defense of Right Reason."

Weatherson, Brian (manuscript). "Do Judgments Screen Evidence?"

Williamson, Timothy (2011). "Improbable Knowing." In T. Dougherty (ed.) *Evidentialism and its Discontents*, Oxford University Press.

Williamson, Timothy (2014). "Very Improbable Knowing." *Erkenntnis*, 79(5): pp. 971–999.