# Epistemic Expansions

## Jennifer Carr

**Abstract:** Epistemic transformations—changes in one's space of entertainable possibilities—are sometimes rational, sometimes irrational. Epistemology should take seriously the possibility of rationally evaluable epistemic transformations. Epistemic decision theory compares belief states in terms of epistemic value. But it's standardly restricted to belief states that don't differ in their conceptual resources. I argue that epistemic decision theory should be expanded to make belief states with differing conceptual resources comparable. I characterize some possible constraints on epistemic utility functions. Traditionally, it's been assumed that the epistemic utility of a total belief state determines the epistemic utility of individual (partial) beliefs in a simple, intuitive way. Naive generalizations of extant accounts generate a kind of repugnant conclusion. I characterize some possible alternatives, reflecting different epistemic norms.

I've never had a child. I've never tasted an oyster. I've never experienced war. I don't have the slightest idea what any of these experiences is like. I can't even entertain the possibilities for what they are like. How can I make rational decisions about whether to have a child, taste an oyster, or go to war, when I have so little idea what kind of outcome I'd generate? Normally I could decide on the basis of thinking about the possible outcomes of my actions: how valuable each would be, and how likely it is. But I can't do that.

This is L. A. Paul's (2015) challenge: what could rationalize a decision about whether to perform any of these actions? How can there be a decision theory for partial credence functions, when decisions hinge on possibilities the agent can't entertain? The problem is not uncertainty: it's not simply that the agent is unsure of the outcomes of her actions. Rather, the problem is limited conceptual resources: there are some possibilities that the agent can't "see," propositions she isn't in a position to entertain.

We can formulate the puzzle, then, in overtly decision theoretic terms. Suppose an agent has only a partial credence function, one that doesn't range over some of the possible outcomes of an action available to her. Then

the expected utility of that action is undefined. So what could rationalize the choice of whether to perform it?

An epistemic analogue to Paul's challenge: how can there be an epistemic decision theory for an agent with a partial credence function, when her epistemic "decisions" hinge on possibilities the agent can't entertain?[1]

Suppose I'm deciding whether to $\phi$, where $\phi$ing involves changing my credences. But I don't know the epistemic utility of $\phi$ing. Its epistemic utility depends on what state of the world I'm in: in particular, on which credence function $\phi$ing will lead me to adopt. What's worse: I can't even entertain these credence functions, because they involve concepts that I don't possess. They distinguish among possibilities I can't distinguish. So how am I supposed to choose whether to $\phi$?

The epistemologized version of Paul's challenge, then, is: can we extend epistemic decision theory for agents with partial credence functions? For agents whose credences can change domain, so that the agent can come to see different possibilities? I'll argue that we can. Moreover, we should: it's sometimes irrational to change the domain of one's credence function.

The plan of the paper is as follows: I argue that epistemic decision theory should be expanded such that credence functions with different domains are sometimes comparable. The argument is based on a week conservative principle: that it's sometimes irrational to lose conceptual resources. Then I characterize some possible constraints on epistemic utility functions that compare credences with different domains. Traditionally, the epistemic utility of a total credence function is understood as a function of the epistemic utility of credences in individual propositions. The most natural ways of generalizing to partial credences generates a kind of repugnant conclusion.[2] I argue for a general constraint on the space of possible algebra-neutral epistemic utility functions. I characterize some possible versions, reflecting different epistemic theories.

# 1 Paul's challenge epistemologized

## 1.1 Partial credence functions

Jackson (1982) argued that before ever seeing the color red, there was something Mary failed to know: what it's like to see red. Jackson's thought experiment arguably supports a stronger conclusion: Mary can't even conceive of what it's like to see red. Of course, she can entertain the possibility that what it's like to see red is the same as what it's like to see dark gray, or the same as what it's like to taste an oyster. But there are many

---

[1] To highlight the analogy with practical decision theory, I use voluntaristic phrasing ("decisions," "options," etc.). But epistemic decision theory doesn't presuppose epistemic voluntarism.

[2] Parfit 1984.

possibilities for what it will be like to see red that Mary can't entertain, including what it's actually like.

Like Paul, we begin with the assumption: people don't always have attitudes toward all propositions. Within the partial belief framework: sometimes a person's credence function isn't defined over all propositions.

This is already controversial. On the traditional view of the psychology of credences, both credence and utility functions are understood as abstractions from dispositions to choice behavior. There are representation theorems that show that, if an agent's dispositions satisfy some constraints, the agent is describable by a unique credence function and a utility function unique up to positive affine transformation. The traditionalist may insist that credence functions aren't partial. Or, if she allows that they may be partial, the traditionalist will insist may insist that at least credence functions are not in any way that could change over time. Credences are automatically defined for all possibilities the agent could possibly encounter, since for each encounterable possibility, the agent will have choice dispositions.[3]

We might respond by simply accepting that the dispositionalist psychology of credences is wrong. But there are also possible replies to this objection that maintain the spirit of the dispositionalist view. First, we might question whether an agent really has the relevant dispositions to choice behavior even for choices that hinge on propositions she can't entertain. It might be that, instead, agents have dispositions to acquire dispositions to choose once the relevant conceptual resources are active.[4]

Second, we might accept that credences are dispositions to choice behavior that meet certain necessary conditions. Plausible theories of intentionality impose some conditions on propositional attitudes. These conditions may be externalist: in order to have *de re* thoughts about individuals, it might be that you need to have a particular sort of causal connection with those individuals. (Shakespeare couldn't have had de re thoughts about Cher.) They may also be internalist: in order to have thoughts about phenomenal redness, it might be that you need to have already experienced phenomenal redness.

A different kind of objection: decision theory characterizes agents who are idealized in all kinds of ways. For example, their credences are infinitely non-vague. Traditional decision theory presupposes credences that are infinitely precise. Fans of imprecise credences functions assume credences take sets of reals as values, and so imprecise credences have infinitely precise boundaries. And so on. Ideally rational agents aren't computationally limited. So why allow that they can have merely partial credence functions?

Reply: if there are external or experiential conditions on having propositional attitudes, then plausibly ideal rationality is compatible with having partial credence functions. Ideal rationality doesn't require us to be in a

---

[3] The traditionalist assumes that agents' dispositions are determinate, though not fixed.

[4] Thanks to Robbie Williams for this suggestion.

certain sort of environment, or to have had certain kinds of perceptual experiences. Mary is conceptually limited not because of some irrationality in her belief state, but because a certain kind of experiential state is a necessary precondition for entertaining what it's like to see red. Rationality doesn't require one to have already experienced red phenomenology.

In short, the presupposition that rational agents have total credence functions is not an intuitive assumption about ideal rationality. It's an idealization only in the sense of "simplification." When this simplification is removed, it opens up substantive questions about rational decision making and rational update.

I'll assume that even partial credence functions are defined over boolean algebras of propositions, the strongest elements of which form a partition over the set of worlds. In other words, I assume there's an exhaustive set of mutually exclusive 'basic possibilities' that are "visible" to an agent; the agent's credences are defined over all unions of visible basic possibilities plus the empty set. What justifies this assumption? Plausibly, while ideal rationality doesn't require being able to see all propositions, it does require being able to negate and conjoin the propositions you do see.

The partition of basic possibilities an agent can see characterizes the distinctions the agent is able to make. As an intuitive shorthand, I'll talk about an agent's "concepts" or "conceptual resources." Note: I don't mean to make any substantive commitments about the psychology of concepts. Indeed, nothing in the way we're modeling things will play the role of "a concept," qua subpropositional mental representation. The distinctions in the space of possibilities that an agent can make are presumably related to her conceptual resources, but I acknowledge that that relation might be very messy. If the shorthand seems misleading or distracting, talk of concepts may be translated into talk of distinctions an agent can make in logical space.

Once we notice that rational agents can't always see all propositions, Paul's challenge arises. How can we make rational choices when we aren't in a position to entertain the possible outcomes of our actions?[5]

## 1.2 Epistemic decision theory

Paul's challenge is a problem for practical decision theory. The epistemologized variant we're considering is a problem for epistemic decision theory.

---

[5] Here, it might be that I'm formulating the challenge differently from how Paul (2015) sees it. On her view, the specification of outcomes won't determine facts about the agent's phenomenology. So an agent whose credences aren't defined over relevant phenomenological propositions will nevertheless be able to entertain the possible outcomes of her acts. But the agent's inability to conceive of the phenomenology of, e.g., having a child will prevent her from assigning utilities to outcomes in which she has a child. This difference may be substantive. Indeed, it may be that by shifting the focus away from phenomenology, what I call "Paul's challenge" would be better called "one of Paul's challenges."

Epistemic decision theory is an application of a Savage-style decision theory, restricting itself to epistemic "acts" and epistemic utilities. Epistemic "acts," typically not construed voluntaristically, involve possessing or coming to possess a credence function. Epistemic utility functions represent comparative epistemic goodness. Commonly defended epistemic decision rules:

> **Dominance:** if credence function $c$ has higher epistemic utility than credence function $c'$ at every world, don't adopt $c'$.
>
> **Maximize Expected Epistemic Utility:** adopt the credence function with the highest expected epistemic utility (i.e. the highest weighted average of epistemic utilities at all possible worlds, weighted by the probability of those worlds).

As with ordinary decision theory, the relevant sort of goodness is ultimate epistemic goodness. Epistemic goodness of this sort is objective, in the sense that it's non-information-dependent. A common view is that the relevant sort of ultimate epistemic goodness should be interpreted in terms of gradational accuracy.[6] Gradational accuracy is the closeness of a credence function to the truth, by some measure satisfying a handful of intuitive constraints. Credence 1 in $p$ is maximally close to the truth iff $p$ is true; credence 0 is maximally close to the truth if $p$ is false.

We can distinguish the epistemic utility of a particular credence an agent has in an individual proposition from the epistemic utility of the agent's total belief state. We'll call the former 'local epistemic utility' and the latter 'global epistemic utility.' It's usually assumed that global epistemic utility is determined straightforwardly as a function of local utilities.

Turning back to Paul's challenge: can epistemic decision theory be extended to credence functions defined over different propositions?

## 2 Epistemic decision theory for partial credences

### 2.1 Motivating comparability

Epistemic decision theory usually presupposes that the credence functions it compares are defined over the same algebra of propositions. Once we abandon this presupposition, new difficulties arise.

For example, to narrow the space of epistemic utility functions, constraints such as 'truth-directedness' and 'immodesty' are placed on epistemic utility functions. Roughly, truth-directedness says credence functions are epistemically better the closer they are to the truth. Immodesty says that probabilistic credence functions assign themselves higher expected epistemic utility than all other credence functions. It's no longer clear what these constraints amount to, or why they are intuitive, once we compare credence functions that are defined over different domains. Should a probabilistic

---

[6] Joyce 1998, 2009; Leitgeb and Pettigrew 2010a,b.

partial credence function assign itself higher expected utility than proba-bilistic extensions of itself that see more propositions? Is the extension automatically closer to the truth than the original?

It's tempting to prescind from these questions. Epistemic decision theory wasn't designed to make comparative evaluations of credence functions that see different propositions. It falls silent about these sorts of comparison. So perhaps credence functions with different domains are incomparable in epistemic value.

Here are two arguments for why credence functions that can see different propositions should be comparable.

First: decision theory aims to be neutral with respect to substantive nor-mative questions. It provides only structural constraints on rational choices. Epistemic decision theory involves placing some substantive constraints on epistemic utility functions, but these are only meant to to delineate the epistemic subject matter and need to be individually justified. There's no intuitive basis for assuming that credence functions with different domains can't be compared. So epistemic utility theory should accommodate the possibility that conceptual change has epistemic (dis)value.

Second: there are intuitive grounds for making at least some credence functions with different domains comparable. Here's an argument from ev-identialism: it's irrational to change your credence in a proposition without new evidence. One way of changing your credence in a proposition is to abandon it, so that your credence function no longer sees the proposition. So when evidence is held fixed, it's irrational to abandon your credence in a proposition. More briefly: it's irrational to undergo an "epistemic contraction." By contrast, it's plausible that it's at least sometimes rational to undergo an "epistemic expansion," whereby you retain your previous credences but come to see new propositions. (Intuitions are murkier about cases where you come to see new propositions and lose sight of old ones, and so I focus narrowly on easy cases: pure epistemic expansions and contractions.)

Let's consider stronger and weaker versions of these claims:

(1) It's always rationally impermissible to undergo an epistemic con-traction.
(2) It's sometimes rationally impermissible to undergo an epistemic contraction.
(3) It's always rationally permissible to undergo an epistemic expan-sion.
(4) It's sometimes rationally permissible to undergo an epistemic ex-pansion.

Of these four claims, all that is needed to establish the possibility of cross-algebra comparison is for one of the weaker claims claim 2 and claim 4 to be true. So here I rely on intuition. There are at least some circumstances where losing a credence in a proposition is irrational. In order for it to be

possible to model the comparison between epistemic states before and after an epistemic expansion or contraction, credence functions with different domains must at least sometimes be comparable.

## 2.2 Strong and weak conceptual conservatism

For the purposes of this paper, I'll commit to claim 2 and claim 4. But in fact, all four have some intuitive plausibility, at least when the space of options is unconstrained. (When the space of options is constrained, there can be cases of forced choices between irrational credences and epistemic contractions, or cases where all the only optional expansions are irrational. In such cases, contractions may be permissible and expansions impermissible, contra claim 1 and claim 3. Note also that my discussion is confined to ideal rationality, where clutter avoidance and other resource constraints are non-issues.)

Consider claim 1, which we can call 'strong conceptual conservatism'. A brief defense of strong conceptual conservatism:

> **Strong evidentialism:** It is irrational to change your credences without acquiring new evidence.
>
> **No evidence against concepts:** There can be no evidence that justifies losing conceptual resources.

Together these entail strong conceptual conservatism.

One might object: aren't there good reasons to abandon some concepts? For example, concepts that have false presuppositions? For example, once I realize that the concept *slut* has false misogynist presuppositions, isn't it best to abandon that concept altogether? In fact, though, it's better to retain the concept.

One might argue that the real objection to having the concept *slut* is a practical, ethical objection, not an epistemic one. Practical norms have no bearing on epistemic norms. (We are not epistemically required to have nonprobabilistic credences even when we know that doing so will magically save the life of a child.)

A better argument for retaining even problematic concepts, on my view, is that doing so is preferable both epistemically and practically. What's really wrong with objectionable concepts is not possessing them, but rather (in some sense) applying them. Suppose you know that Joe believes that Mary is a slut. If you abandon the concept of a *slut*, then you will no longer know exactly what Joe believes about Mary. If you want to talk him out of this misogynist belief, better to keep the concept *slut* and explain to him what's wrong with applying it.

Nothing in the present framework models a fully general distinction between possessing and applying a concept, but the idea should be roughly clear. Refraining from applying a concept might mean rejecting (having

credence 0) in all propositions that apply the concept. (One might wonder: how is that consistent with Bayesianism? Wouldn't you have to have credence 0 in both a proposition and its negation, thereby violating probabilism? Response: to make sense of concepts with false presuppositions in this sense, we'd have to move to a nonclassical setting.)[7]

So: problematic concepts don't generate a problem for strong conceptual conservatism. But there are other possible different angles of attack.

First, it might be that in some cases, if you face a choice between having credences that employ a problematic concept in an irrational or otherwise problematic way, or else losing the concept altogether, then perhaps you should choose the latter. For example, suppose it's psychologically impossible for you to retain the concept *slut* without lending positive credence to propositions that apply it. Then perhaps it's better just to lose the concept.

Here I think, the fan of strong conceptual conservatism will reply: if your credences are already irrational, then rationality doesn't recommend retaining them. Contracting your conceptual resources is still irrational, but it might be the lesser of two epistemic evils. What this objection is really an objection to is the stronger claim that in all cases, it's better to retain one's own credence function than to adopt any contraction of it. That is plausibly false, but not something that the strong conceptual conservative is committed to.[8]

A different type of objection to strong conceptual conservatism suggests that sometimes evidential considerations favor loss of concepts. For example, it might be that as our knowledge of the world develops, our concepts change; a naive concept is replaced by a more sophisticated one, or two separate concepts merge when they're revealed to be extensionally equivalent or analytically identical.

But these sorts of objections rely on a more inflated theory of conceptual resources than capacities to draw distinctions in logical space. Extensionally equivalent and analytically identical concepts don't draw different distinctions in logical space. Concept change over time may involve getting rid of concepts in some psychological sense, but it's not clear that it rationally requires abandoning distinctions in logical space.

---

[7] Robbie Williams has pointed out to me that, in the case of concepts with false presuppositions, the problem may be more complex than I make out. On some views, each concept with a false presupposition determines a concept with the same application conditions that doesn't trigger the false presupposition; and so one can abandon the false presupposition while retaining the same distinctions in logical space by switching to the equivalent, presuppositionally innocuous concept. This suffices for conceptual conservation in the sense of "conceptual" I use throughout. But in the case of thick concepts like *slut*, this form of disentanglement may not be possible, and so other resources (e.g. rational but nonprobabilistic credences) may be necessary. With still other forms of problematic concept, we may have to tell a different story. Still, for the purposes of representing others' beliefs, or merely possible beliefs, or counterfactual or counterpossible scenarios, I suggest, strong conceptual conservatism is compelling.

[8] Thanks to L. A. Paul, Julia Staffel, and Robbie Williams for discussion.

So these observations are in fact compatible with conceptual conservatism, in my (loose, perhaps unfortunate) sense of the word "conceptual." There are, of course, interesting questions about the epistemic value of concepts in a more inflationary sense; but that's not under discussion here.

A final objection to strong conceptual conservatism: it prohibits ever losing conceptual resources. This is a diachronic epistemic constraint. It's controversial whether there are diachronic constraints on rationality.[9]

Opponents of diachronic constraints on rationality reject strong evidentialism, in favor of a weaker variant:

> **Weak evidentialism:** It's irrational to change your credences without some change in evidence.

Weak evidentialism allows changes in your credences if you either gain new evidence or lose old evidence, e.g. by forgetting information. On its own, weak evidentialism entails claim 2, which we can call 'weak conceptual conservatism.' Indeed, it entails something stronger:

> 2′ 'Medium-strength conceptual conservatism': it's rationally impermissible to undergo an epistemic contraction without some change in evidence.

We can equally provide arguments from claim 4 to comparability. For example, suppose $c$ is a partial credence function, and $c^+$ is an extension of $c$ such that for all propositions $A$ visible to $c^+$ and not $c$, $c^+(A)$ is maximally accurate (or otherwise has maximal epistemic value). Intuitively, $c^+$ must be as accurate as (or as valuable as) $c$. And so $c$ and $c^+$ must be comparable.[10]

The upshot: evidentialism supports the conclusion that at least some credence functions with different domains can be compared.

## 2.3 Impact on epistemic decision theory

What this means for epistemic decision theory: we need epistemic utilities, pr epistemic decision rules, that make at least some credence functions defined over different domains comparable.

One natural temptation would be to be as neutral as possible with respect to the epistemic utilities of partial credence functions: for example, to assign partial credence functions only imprecise epistemic utilities. A neutral imprecise utility assignment for a partial credence function $c$ would equal the set of global utilities of all total extensions of $c$.

How do we compare imprecise utilities? Perhaps the imprecise utility of $c$ will be greater than the imprecise utility of $c'$ iff the precise utilities of all total extensions of $c$ were greater than the precise utilities of all total

---

[9] Against diachronic rationality, see Talbott 1991, Christensen 2000, Williamson 2000, Meacham 2010, Moss Unpublished, Hedden 2013.

[10] Thanks to an anonymous referee for suggesting this compelling intuition pump.

extensions of $c'$. The problem is that this generates widespread incompara-
bility. In particular, partial credences will always be incomparable with all
of their extensions. The imprecise utility of a credence function will be a
superset of the imprecise utility of its extensions.[11]

We may be happy not to compare the utilities of credence functions
defined over disjoint sets of propositions, or overlapping sets of proposi-
tions where neither includes the other. But comparing credence functions
and their extensions was supposed to be the easy case. An algebra-neutral
epistemic utility function should be able to compare at least some par-
tial credence functions and their extensions. Otherwise we can't predict
conceptual conservatism.

It would be hasty to rule out imprecise utilities altogether. But I've
argued that epistemic decision theory should allow at least some credence
functions over different domains to be comparable, including at least some
credence functions and their extensions. So either way, we can't avoid
substantive epistemological questions about what constraints there are on
epistemic utility functions that range over credence functions with different
domains.

## 3 Epistemic utility functions for partial credences

Instead of retaining a utility function that ranges only over total credence
functions, and assigning partial credence functions imprecise utilities, I
suggest we consider epistemic utility functions that range over (at least
some) credence functions with different domains.

Which utility functions? Instead of positing unique epistemic utility func-
tions, epistemic utility theory generally proceeds by imposing constraints on
the space of candidate epistemic utility functions, constraints which suffice
represent epistemic norms of various sorts. We can think of principles like
strong and weak conceptual conservatism as constraints on the space of
epistemic credence functions, constraints that encode the epistemic value of
conceptual resources. There may be other intuitive principles linking the
utilities of partial credence functions and their extensions.

Other constraints on algebra-neutral epistemic utility functions may be
natural generalizations of accepted constraints on algebra-specific functions.
For example, in place of strict propriety (the constraint that algebra-specific
utility functions should make probabilistic credence functions assign them-
selves higher expected utility than all alternatives), an algebra-neutral

---

[11] Obviously, we might use a different rule for comparing imprecise credences: for example,
perhaps $c$ is strictly preferable to $c'$ iff the maximal utility in $c$ is greater than the maximal
utility in $c'$ and the minimal utility in $c$ is greater than the minimal utility in $c'$. Then we
might generalize: $c$ is weakly preferable to $c$ iff the maximum utility of $c$ is greater than or
equal to the maximum utility of $c'$ and the minimum utility of $c$ is greater than or equal to
the minimum utility of $c'$. Even then, we can never predict that an extension of $c$ is strictly
preferable to $c$, and so we can't predict weak conceptual conservatism.

epistemic utility function should perhaps satisfy a generalization of strict propriety that makes probabilistic partial credence functions assign themselves higher expected utility than all alternatives defined over the same partition.

### 3.1  Local and global utilities

In order to extend epistemic decision theory to partial credence functions, we need to look at how adding or subtracting individual credences affects the epistemic utility of an agent's total epistemic state. In other words, we need to look at the relationship between the local utilities of credences in individual propositions and the global utilities of overall credence functions.

Global epistemic utility is usually interpreted (noncommittally) as a sum or average of local epistemic utilities. Where $g$ and $l$ are global and local utility functions, respectively, and $\mathscr{A}$ is the algebra of propositions $c$ is defined over,

**Summing proposal:**
$$g(c,w) = \sum_{A \in \mathscr{A}} l(c,A,w)$$

**Averaging proposal:**
$$g(c,w) = \frac{1}{|\mathscr{A}|} \sum_{A \in \mathscr{A}} l(c,A,w)$$

Now, epistemic utility is standardly understood as gradational accuracy. Accuracy-first epistemic decision theory assigns credences value in terms of their distance from the truth, by some measure satisfying a handful of intuitively plausible constraints.

So we can ask: is an accuracy-first epistemology compatible with treating conceptual resources as epistemically valuable? The answer can be *yes* only if the epistemic value of conceptual resources is reducible to the epistemic value of accuracy. As it turns out, both the summing hypothesis and the averaging hypothesis for global (in)accuracy introduce commitments about the value of conceptual resources. These commitments potentially are unattractive.

We could, of course, switch to epistemic utility functions that aren't accuracy-directed. Indeed, we might be forced to do so in order to avoid implausible conclusions about the value of conceptual resources. But it's worth exploring whether we can represent the value of conceptual resources without departing to far from accuracy-first epistemology, especially because the research project has otherwise proven fruitful.

### 3.2  Summing positive disutilities

In accuracy-first epistemic decision theory, local inaccuracy is interpreted as a positive penalty for distance from the truth. There is a maximum degree

of accuracy, which is distance 0 from the truth. Any credence other than 1 in truths or 0 in falsehoods incurs a positive disutility.

And so summing local disutilities generates an immediate consequence: if $c^+$ is an extension of $c$ and any of the new credences it brings are uncertain, $c^+$ will incur whatever disutilities $c$ has plus disutilities for its new credences. So $c^+$ will automatically have greater global disutility $c$. In other words, $c$ dominates $c^+$, merely because $c^+$ can make new distinctions but isn't omniscient about them.

So on this proposal, we find ourselves with the result that the fewer imperfect credences you have, the better you are epistemically. This amounts to treating nonattitudes toward propositions as epistemically *perfect*: maximally accurate. They are matched only by the epistemic utility of credence 1 in truths and credence 0 in falsehoods.

This proposal, paired with either weak dominance avoidance or expected utility maximization, yields the following verdicts: rational agents only have credences in propositions such that their credence matches the truth value at every world (i.e. ⊤ and ⊥ and any other known necessities). Specifically, rational agents will have credence 1 in all tautologies, credence 0 in all contradictions, and no other defined credences. And of course, similar problems afflict any other, non-accuracy-based interpretations of global epistemic disutility that treat global disutilities as sums of positive disutilities.

I take it as a datum that epistemic utility functions shouldn't universally prohibit attitudes toward contingent propositions, or propositions the agent can't be certain of.

## 3.3 Averaging

What if instead of treating the global (dis)utility of a credence function as the sum of its local (dis)utilities, we treated it as the average?

Then there is no automatic dominance relation between credence functions and all of their extensions. Sometimes seeing new propositions will increase global utility, sometimes decrease it. At some worlds, new uncertain credences may increase your average accuracy; at others decrease it.

The averaging proposal seems like an obvious move. But it brings with it new problems. Suppose you have credence function $c$, which sees propositions $A$ and $\neg A$ and assigns .8 in $A$, which is true. Then suppose you have the option of extending your credence function to $c^+$, which also sees new propositions $B$ and $\neg B$, and has credence .6 in the true one.[12]

Two consequences. First, adopting $c^+$ entails a reduction in your global accuracy at the actual world. Your old credences in $A$ and $\neg A$ were pretty accurate. Your new credences in $B$ and $\neg B$ are on the right track, but

---

[12] To keep the example as simple as possible, I don't stick to my assumption that credences must defined be over the boolean closure of a partition.

they're still not as close to the truth as your old credences were. So they drag the average down.

Second, adopting $c^+$ will typically entail a reduction in your expected global accuracy. After all, you are more confident than not that you're at a world where your average accuracy is dragged down.[13]

Two objections. First, it's not clear that this is an intuitive way of characterizing the overall accuracy of a credence function. Second, if we accept this as a characterization of the accuracy of a credence function, then accuracy is a bad measure of epistemic value.

On the first point: it's clear that your average local accuracy is reduced when you move from $c$ to $c^+$. But it's not obvious that your global accuracy should be reduced. There is some intuitive sense in which $c^+$ is doing better at the actual world, accuracywise, than $c$. For one thing, $c$ couldn't distinguish $B$ from $\neg B$. $c^+$ not only distinguishes them, but is closer to truth than to falsehood. So this example may motivate rejecting the equivalence of global accuracy with average local accuracy.

On the second, less controversial point: if global accuracy is average local accuracy, then it's implausible that global accuracy is a good characterization of epistemic value.

In this example, suppose our agent is an expected utility maximizer and epistemic utility is simply average accuracy, measured by the most familiar accuracy measure, the Brier score.[14] Then if she has $c^+$ as her credence, she'll prefer to abandon her attitudes toward $B$ and its negation. $c$ has higher expected accuracy from $c^+$'s perspective than $c^+$ itself.

The problem doesn't just afflict expected utility maximizers. Even restricting ourselves to strong accuracy dominance avoidance, we end up at the bottom of the same slippery slope. If the agent has attitudes towards propositions that she's uncertain of, then her average accuracy is imperfect. But if she only has attitudes toward propositions that she's certain of— again, tautologies and contradictions—then her attitudes will have perfect average accuracy, and hence will be more accurate at every possible world. So in order to avoid having credences that are strongly dominated, she must restrict her credences to propositions that she's certain of. In order to avoid having credences that are weakly dominated, she must restrict her credences to propositions that are necessary across all possibilities that she can entertain.

---

[13] This depends on the choice of local utility function or accuracy measure, since the loss of average accuracy at some worlds may be offset by a more dramatic gain in average accuracy at other worlds. But for any continuous, truth-directed accuracy measure, we can generate an analogous case where the expected average accuracy will decrease with the addition of new credences.

[14] The Brier score of a probability is the squared Euclidean distance between the probability assigned to a proposition and its truth value (1 = true and 0 = false).

And so, like the summing disutilities proposal, I think we have good reason to reject a algebra-neutral epistemic utility function that derives global utilities from averages of local utilities.

### 3.4 Summing positive utilities

Let's return to the summing proposal. What if instead of penalizing distance from truth, we reward distance from falsehood? That is, what if we derive global utility by summing positive local utilities? Then, of course, the situation is reversed with respect to summing positive disutilities. Each new proposition added to the domain of a credence function increases the credence function's epistemic value, as long as the credence it assigns isn't maximally inaccurate. This means treating nonattitudes towards propositions as maximally inaccurate—just as bad, from an epistemic point of view, as certainty of falsehood.

This is not counterintuitive in the way that summing disutilities is. There's a perfectly reasonable position in logical space according to which any increase in conceptual resources contributes positively to epistemic utility. It says that conceptual resources are a pure epistemic good: they trump any epistemic badness that might be required to achieve them. So it automatically gains us strong conceptual conservatism.

Whereas summing disutilities generated a dominance argument against having credences in any contingent propositions, summing positive utilities generates a dominance argument for having credences in every proposition. One might object: doesn't this contradict the premise of this paper: that there can be rational agents with partial credence functions? Doesn't this entail that before she sees red, Mary is irrational?

Answer: having dominated credences is only irrational if there are any non-dominated credences that are among one's epistemic options. In Mary's case, though, adopting credence functions that can conceive of what it's like to see red is not an option for her. She would need to experience phenomenal redness before she'd be in a position to adopt that credence function. But she can't: she's locked in a black room. So she's not irrational for having a credence function that's dominated only by a non-option, any more than you are irrational for not spontaneously acquiring a billion dollars.

It does mean, though, that if your credence function is needlessly partial, then you're irrational. If you simply fail to have credences in some possibilities, even though you're perfectly well in a position to do so, then this view judges you irrational. Two reasons to think this is not a problem: first, it's presumably controversial whether this is even psychologically possible. Second, if it is, it's not obvious that it should be rational.

A different kind of objection: suppose there were a way to make up new concepts. For example, you might give your friend Rachel a separate name, "Srachel," just for Tuesdays. Then you might conceptually distinguish

Srachel (who went to karaoke on Tuesday) from Rachel (who gave a talk on Wednesday). Even though you know they're the same person, perhaps you can now conceive of the possibility that they're two different people.

The details of the example don't really matter; we simply need some way for a person to be able to generate new concepts. Then, the proposal seems to suggest that doing so is epistemically mandatory. That's implausible.

First, it's not clear that the act of voluntarily inventing a new concept is an epistemic act, as opposed to a practical one. Second, it's not clear that in introducing new concepts of this sort, you're really exposing any distinctions in logical space. (After all, you plausibly already have the concept of *rachel on tuesdays*; possessing that concept was instrumental in your invention of the *scrachel* concept. Third, if somehow you are exposing real distinctions in logical space, then it's perhaps doing so is epistemically preferable after all. The only obvious reason against it is clutter avoidance; but for ideally rational agents I take it that epistemic clutter is unproblematic.

Summing positive utilities does have some questionable consequences, though. For example, suppose a partial credence function $c$ has an extension, $c^+$. Then $c^+$ dominates $c$, regardless of what credences it assigns to new propositions. $c^+$ could be irrational in lots of ways: it might assign credence 1 to both a proposition and its negation. Still, $c^+$ would dominate $c$, even if $c$ were probabilistically unimpeachable.

This may not be a deep problem. If $c^+$ isn't probabilistic, then it will itself be dominated by some probabilistic credence function defined over the same algebra.[15] And so as long as an agent's epistemic options include all those credence functions defined over subsets of the propositions potentially visible to the agent, the summing utilities proposal will never require one to adopt nonprobabilistic credences.

Still: it's not obvious that seeing new propositions in an irrational way is necessarily epistemically better than not seeing them at all, even when both are suboptimal.

There are other reasons why one might not be satisfied with this proposal. Because it treats nonattitudes toward propositions as having the lowest possible epistemic utility, expanding one's epistemic vision is an absolute epistemic good. One might doubt this: for example, one might think that having a very inaccurate credence in a proposition is worse, at a world, than not having any credence in that proposition.

I don't want to rule out the summing positive utilities proposal. It has many intuitively appealing features, including its preservation of conceptual conservatism. But these considerations suggest that it's worth exploring other options. There's room in first-order epistemology for controversy

---

[15] Assuming our algebra-neutral utility functions preserve the probabilism-entailing properties of algebra-specific epistemic utility functions when comparing credence functions that share an algebra.

about whether new conceptual resources are always an absolute epistemic good, or whether their utility can be outweighed.

### 3.5  "Better than chance"?

So far, a pattern has emerged. When we sum positive disutilities, in effect we treat nonattitudes toward propositions as if they had maximal epistemic utility. When we sum positive utilities, in effect we treat nonattitudes toward propositions as if they had maximal epistemic disutility. When we average local (dis)utilities, in effect we treat nonattitudes as if they had the same utility as the average utility of attitudes.

Instead of treating nonattitudes as though they all having maximal utility or minimal utility, it seems plausible that we should treat nonattitudes as having middling utility. The averaging proposal accomplishes that, but not in the right way.

One might be tempted to say: nonattitudes toward propositions are no closer to truth than to falsehood. And so they should have the same utility as credences that are no closer to truth than falsehood, i.e. credence .5.

Something like this might be on the right track. But suppose Mary has no attitude toward the propositions that seeing red is like experiencing phenomenal redness, or like experiencing phenomenal blueness, or like experiencing phenomenal azureness, or . . . . Suppose Mary's partial credence function has the same utility as an extension of itself that assigns credence .5 in *all* of these propositions. Then it has the same utility as a wildly non-probabilistic credence function. This will generate unpredictable consequences for when a very inaccurate credence function has higher utility than locally very accurate but partial credence function. In any case, this certainly makes partial credence functions dominated (since all non-probabilistic credence functions are dominated, assuming familiar constraints on epistemic utility functions). This is a substantive, non-obvious epistemological hypothesis.[16]

To avoid this problem, the local utility of nonattitudes need not be the same for both truth and falsehood. That does satisfy a plausible desideratum: that the global utility of a partial credence function is uniform across worlds it doesn't distinguish. But the proposal is stronger than necessary: one might satisfy the same desideratum without assuming that a nonattitude's local utility is the same at worlds where it's true and worlds where it's false. Local utilities may differ as long as it all evens out at the global level. Let's consider another sort of proposal that generates uniformity in global utilities across indistinguishable worlds.

Consider again the example in the last subsection. A fan of accuracy-first epistemology might reason: even though $c^+$ brings down average local accuracy, it seems to be doing pretty well, accuracywise. After all, it assigns .6 credence in $B$, a truth; so it's closer to truth than falsehood. One might

---

[16] Thanks to Kenny Easwaran for prompting me to consider this proposal and for discussion.

say, intuitively, $c^+$ is doing "better than chance". Maybe that's the sense in which it seems to be an improvement in accuracy over having a nonattitude.

And so, one might suppose nonattitudes toward propositions are worse than credences that are more accurate than chance, but better than credences that are less accurate than chance. The credences that are neither worse nor better than chance are the maximally unopinionated attitudes. So, on this view, nonattitudes have the same utility as maximally unopinionated attitudes.

Spelling this out: suppose each partial credence function $c$ has a unique maximally unopinionated total extension: call this credence function $c^{\dashv}$. On the hypothesis we're considering, the utility of $c$'s nonattitude toward $A$ is equal to the utility of $c^{\dashv}$'s attitude toward $A$.

This hypothesis has a number of attractive features. Unlike summing disutilities or averaging, there is no automatic epistemic gain in seeing fewer propositions. There is also no automatic epistemic gain in seeing more propositions: this view allows that whatever epistemic good there is in seeing new propositions, it can be outweighed.

Another feature of this view is that it ensures weak conceptual conservatism, and comes close to ensuring strong conceptual conservatism. If $c$ is probabilistic, then so is $c^{\dashv}$. And so if $c$'s utility matches $c^{\dashv}$'s, then with standard (strictly proper) scoring rules, $c$ will be non-dominated and will maximize expected utility from its own perspective. So it will never be rationally required to give up having any credence at all in some propositions. And, if the agent is an expected utility maximizer, she'll prefer her own partial credence function to almost any alternative the domain of which is a subset of hers.[17]

But this hypothesis does face standard symmetry worries. It depends on the existence of a unique, maximally unopinionated extension of $c$. But problems for the principle of indifference suggest that there isn't any objective basis for determining a unique maximally unopinionated $c^{\dashv}$.[18] In other words, there is no objective way of isolating a uniform distribution of probabilities $c_0$ over $\mathscr{W}$ such that $c^{\dashv}$ is (roughly) $c_0$ updated on $c$'s evidence.

This is obviously a legitimate worry. And it isn't easy to avoid: for this style of proposal, we need maximal unopinionatedness to ensure that the global utility of $c$ is uniform over worlds that $c$ can't distinguish.[19] It is

---

[17] It will, however, be epistemically permissible for an agent to give up credence in a proposition if her credence is already maximally unopinionated (and if giving up that credence doesn't violate the requirement that credences be defined over a boolean algebra).

[18] See van Fraassen 1989.

[19] This rules out a natural generalization. Objective Bayesianism—the view that there's a unique rational prior credence function—was originally envisioned as recommending absolute unopinionatedness, i.e. the uniform distribution. In the face of symmetry worries, contemporary objective Bayesians typically think that the rationally privileged prior need not be the uniform distribution over worlds. Suppose we said that the utility of a nonattitude toward a proposition at a world is equal to the utility of the credence assigned by the rationally privileged prior, updated on the partial credence holder's evidence. If the prior is not uniform,

plausibly a desideratum of an algebra-neutral epistemic utility functions that if a partial credence function doesn't distinguish two worlds, then the credence function should have the same epistemic utility at both worlds.

If we like this proposal, we might have to accept that there's some arbitrariness or assessment-sensitivity in the assignment of epistemic utilities. The motivating thought behind this proposal: having a credence in a proposition is better (at a world) than having no attitude toward it if the credence assigned is "better than chance": if one's credence is closer to the truth than if one were to withhold judgment as much as possible. What amounts to doing "better than chance" depends here on a conception of chance as a uniform distribution over indistinguishable worlds, and so presupposes a specific space of worlds. The relevant space of worlds depends on the perspective of the theorist.

## 4 Conclusion

We've seen a variety of accuracy-based proposals for algebra-neutral epistemic utility functions. The most natural extensions of traditional epistemic utility functions—summing local disutilities and averaging local (dis)utilities—both generate terrible consequences. They require agents to give up any credence in propositions that the agent can't be certain of. And so they violate both strong and weak conceptual conservatism with respect to epistemically contingent propositions.

The other two proposals we've seen don't face this problem. Summing positive local utilities entails strong conceptual conservatism. Any loss of concepts will generate a loss of epistemic utility and so will be dominated. And the final proposal—where partial credence functions' utilities match that of their maximally unopinionated extensions—entails at least weak conceptual conservatism.

It's obvious that my discussion has been far from exhaustive. There may be other plausible ways of generating algebra-neutral epistemic utility functions that still make epistemic utility a function of accuracy. There are certainly other plausible epistemic utility functions that don't rely on accuracy alone. It is not the ambition of this paper to narrow down the space of epistemic utility functions to one or the other of these proposals.

What I hope to have made clear, however, is the need for certain constraints on algebra-neutral epistemic utility functions. First, they must permit rational agents to have some uncertainty, and to have attitudes toward contingent propositions. Second, they must permit rational agents to have attitudes toward contingent propositions. Neither of the traditional

---

though, then it might assign different credences to worlds that the partial credence function under evaluation can't distinguish. And so its utility at those worlds might differ.

ways of aggregating local inaccuracies have satisfied these constraints. Furthermore, there is some intuitive support for stronger constraints: perhaps weak conceptual conservatism; perhaps strong conceptual conservatism.

Jennifer Carr

References:

Christensen, David. 2000. "Diachronic Coherence Versus Epistemic Impartiality." *Philosophical Review* 109 (3): 349–371.

Hedden, Brian. 2013. "Options and Diachronic Tragedy." *Philosophy and Phenomenological Research* 87 (1).

Jackson, Frank. 1982. "Epiphenomenal Qualia." *The Philosophical Quarterly* 32: 127–136. http://dx.doi.org/10.2307/2960077.

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

Joyce, James M. 2009. "Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief." In *Degrees of Belief*, edited by Franz Huber and Christoph Schmidt-Petri, Vol. 342, 263–297. Dordrecht: Springer.

Leitgeb, Hannes and Richard Pettigrew. 2010a. "An Objective Justification of Bayesianism I: Measuring Inaccuracy." *Philosophy of Science* 77 (2): 201–235.

Leitgeb, Hannes and Richard Pettigrew. 2010b. "An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy." *Philosophy of Science* 77 (2): 236–272.

Meacham, Christopher J. G. 2010. "Unravelling the Tangled Web: Continuity, Internalism, Non-Uniqueness and Self-Locating Beliefs." *Oxford Studies in Epistemology* 3: 86–125.

Moss, Sarah. Unpublished. "Credal Dilemmas."

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

Paul, L. A. 2015. "What You Can't Expect When You're Expecting." *Res Philosophica* 92 (2).

Talbott, W. J. 1991. "Two Principles of Bayesian Epistemology." *Philosophical Studies* 62 (2): 135–150.

van Fraassen, Bas. 1989. *Laws and Symmetry*. Oxford: Oxford University Press.

Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.