Canadian Journal of Philosophy (2025), 1–12
doi:10.1017/can.2025.10015

ARTICLE

Humility and Complexity

Daniel Greco

Department of Philosophy, Yale University, New Haven, CT, USA
Email: daniel.greco@yale.edu

Abstract

To what extent can intellectual humility be formalized? One natural idea links humility to open-mindedness, captured by a regularity principle: no coherent hypothesis should get probability zero. While debates over regularity often concern infinities, my objection is different. Regularity is feasible only for ideally rational, logically omniscient agents. Yet on a common view, humility involves appreciating our limitations—including our failure to be such agents. So whatever its merits for ideal cognition, regularity is a poor model for human humility. Indeed, taking it as such would itself be un-humble, by failing to appreciate our own epistemic limitations.

Keywords: humility; epistemology; probability; bayesian; idealization

1. Introduction

Epistemology concerns topics like rationality, evidence, knowledge, and related notions. Formal epistemology approaches these topics using formal tools, such as those of probability theory and logic. Much recent work in epistemology concerns the relationship between formal and informal approaches to the subject matter. For instance, what connections are there between the formal epistemologist's notion of degree of belief or subjective probability and the informal, pretheoretical notions of belief and knowledge?[1] What lessons concerning the epistemic norms that apply to computationally limited creatures like us can be learned by studying the idealized, perfectly rational agents characteristic of formal epistemology?[2]

My aim in this article will be to offer a provisional, pessimistic judgment concerning one attempt at bridge-building. Intellectual humility has been much discussed in recent informal epistemology.[3] And one important aspect of humility—open-mindedness—might seem like a promising target for a (partial) formal regimentation. We might think an open-minded agent does not completely rule out coherent hypotheses. Failing to rule out a hypothesis is naturally represented as assigning it some positive probability, so it's natural to suggest that the informal idea of a humble, open-minded agent corresponds to the formal idea of a probability function that assigns some probability $n > 0$ to each coherent hypothesis H. Ultimately, I'll argue against this suggestion.

Here's how the article will be structured. In Section 1, I'll introduce a popular account of intellectual humility (Whitcomb et al., 2017) and will motivate the idea that open-mindedness is plausibly an important part of intellectual humility, so understood. I'll also review some influential arguments that rational agents are open-minded or "regular" in the formal sense that

[1]For exmaple, see Moss (2013), Leitgeb (2014), Moon (2017), Jackson (2020), Sturgeon (2020).
[2]See Yap (2014), Appiah (2017), Carr (2021), Greco (2023a,b), Thorstad (2024), Zhang (2024) for some examples.
[3]See Snow (2018) for a helpful survey.

they do not assign probability 0 to any coherent hypotheses. I'll go on to discuss the literature in formal epistemology on regularity, largely to distinguish extant objections from the objection that I'll ultimately offer. In Section 2, I'll review an influential objection not to regularity per se, but to the Bayesian, probabilistic framework for representing rational belief more generally. As the objection goes, making extensive use of probabilistic reasoning is intractably complicated for computationally limited agents like us, so formal frameworks in which beliefs are represented as probabilities and learning is represented as Bayesian conditionalization have few descriptive or normative lessons for human cognition. I'll argue that this line of thought is better understood as an objection to a particular (popular) interpretation of the Bayesian framework, which requires regularity, rather than to the framework itself, which does not. Finally, in Section 3, I'll sketch a two-stage model of cognition in which a rational agent facing a decision or inference problem first constructs a "small-world" to analyze and then analyzes it using probabilistic tools. I'll explain how cognition so understood can make extensive use of probabilities while not being intractably complicated and violating regularity where it *does* use probabilities. In this sketch, the distinction between appropriately open-minded agents and dogmatically close-minded ones shows up at the hard to formalize the first stage of cognition, rather than the second stage, where probabilities are in play.

2. Humility

According to Whitcomb et al. (2017), humility in general is having the right stance toward one's limitations. Intellectual humility in particular, then, is having the right stance toward one's intellectual limitations (p. 8). What *is* the right stance? It depends on the limitations, but in general, having the right stance toward one's intellectual limitations will involve both understanding what those limitations are and calibrating one's beliefs and behavior in light of them. One intellectual limitation common to all humans is fallibility. As Descartes reminded us, there are few, if any, topics or questions on which we cannot inadvertently slide into an error. What's involved in taking the proper stance toward our fallibility? In Whitcomb et al.'s terms, what does it take to "own" our fallibility? A natural thought is that owning one's fallibility is incompatible with being certain in any fallibly formed beliefs. If my eyesight is fallible, then I should not be certain my eyes aren't deceiving me. If my powers of inference and deduction are fallible, then I should not be certain that some conclusion I've drawn using those powers is correct. Owning one's fallibility, then, requires open-mindedness about the possibility that one has made a mistake.

And if fallibility is ubiquitous, then there will be few, if any, topics about which we can be certain, compatible with our being intellectually humble. Christensen (2007) offers an argument very much along these lines, though he puts things in terms of a tension between certainty and rationality, rather than certainty and intellectual humility:

> Suppose I've never made a mistake in balancing my checkbook or in any other demonstrative reasoning. Surely that does not license me in being certain that such mistakes are impossible. And as long as such mistakes are possible, it is hard to see how I can be certain that they have not occurred. Even if my reason for doubt is slight, and, so to speak, metaphysical—so slight that in ordinary cases, I would not bother to think about it—still, it would seem irrational to be absolutely certain that I had not come to believe a false claim due to a cognitive mistake. And thus it would seem irrational for me to be absolutely certain of [the conclusion of some demonstrative reasoning].

A closely related argument that intellectual humility—or rationality, or reasonableness, or open-mindedness—is incompatible with certainty appeals to the need to make room for learning. Roughly, the idea is that if you are certain that some hypothesis is false, then you will not be in a position to recognize anything as evidence in favor of it. But any coherent hypothesis *could* turn out

https://doi.org/10.1017/can.2025.10015 Published online by Cambridge University Press

to gain support from one's future evidence, so one should not rule it out to start with. Lewis (1980) gives a pure version of this argument:

> Let C be any reasonable initial credence function…I should like to assume that it makes sense to condition on any but the empty proposition. Therefore I require that C is *regular*: $C(B)$ is 0, and $C(A|B)$ is undefined, only if B is the empty proposition, true at no worlds… The assumption that C is regular will prove convenient, but it is not justified only as a convenience. Also it is required as a condition of reasonableness: one who started out with an irregular credence function (and who then learned from experience by conditionalizing) would stubbornly refuse to believe some propositions no matter what the evidence in their favor (Lewis, 1980, p. 271).[4]

And Carroll (2017) offers a sort of hybrid of the argument from fallibility, and the argument from learning:

> … your degree of belief in an idea should never go all the way to either zero or one. It's never absolutely impossible to gather a certain bit of data, no matter what the truth is—even the most rigorous scientific experiment is prone to errors, and most of our daily data-collecting is far from rigorous. That's why science never "proves" anything; we just increase our credences in certain ideas until they are almost (but never exactly) 100%. Bayes's Theorem reminds us that we should always be open to changing our minds in the face of new information… (2017).

These arguments can seem pretty compelling. What resistance have they met with in the philosophical literature? By and large, objections have concerned difficulties with infinity.[5] To use a well-worn example, imagine a fair lottery on the natural numbers—one number will be picked, and each number has an equal chance of being picked. For any given number, you cannot be sure it will not be picked. You can imagine learning that it was picked, after all. So by the arguments above, your degree of belief that some number, say 100, is picked but cannot be 0. But if this means your degree of belief that 100 is picked must be some positive number, then we quickly face a paradox. Either you make invidious distinctions among numbers that by stipulation have the same chance of being picked[6] or you assign each number the same positive probability k of being picked, and end up with a probability that *some* number is picked which is, incoherently, greater than 1. After all, however, small k is, there's some finite m such that $km > 1$.

Advocates of regularity who discuss this problem, such as Lewis (1980, p. 271), accordingly tend to appeal to infinitesimal probabilities—probabilities that are greater than 0, but smaller than any real number k. They then hold that in a fair lottery on the natural numbers, each number has an infinitesimal probability of being picked. Their opponents argue that even the appeal to

[4]While Lewis explicitly restricts his endorsement of regularity to *initial* credence functions—he allows that rational credence functions will become irregular as subjects learn from experience—most writers who endorse versions of regularity take it to have broader scope. And I think Lewis' restriction is hard to motivate. Here's one coherent possibility: I become certain of a proposition that was actually false. By Lewis' lights, it's hard to see how a rational agent can ever get evidence that will make her confident of this possibility; after all, if an agent updates by conditionalization on E, and is now irregular, she will not be able to recognize any future evidence against E. But from her initial, a priori standpoint, it's at least possible that she should make a mistake in becoming certain that E is true. If she thinks, ex ante, she might mistakenly become certain that E is true, it's hard to see why, ex post, she should feel comfortable completely excluding the possibility that her confidence in E—here, E stands in for any proposition she's learned—is misplaced. But then the same learning argument that Lewis uses in defence of regularity for *initial* credence functions can be extended to apply even to noninitial credence functions.

[5]See Easwaran et al. (2021), section 6.3) for a helpful survey, especially the supplement on God's lottery.

[6]By "invidious distinctions," I refer to probability assignments that decay quickly enough so that the whole space can have a probability of 1, for example, you could assign 1 a probability of $\frac{1}{2}$ being picked, 2 a probability of $\frac{1}{4}$, and so on.

https://doi.org/10.1017/can.2025.10015 Published online by Cambridge University Press

infinitesimal probabilities cannot save the defender of regularity from the paradox.[7] My own sympathies are with the opponents of regularity; I do not see any way to "quarantine" the difficulties with infinities that clearly beset straightforward statements of a regularity principle. But I also recognize that many readers will find this line of objection unconvincing. Infinities are weird! They lead to paradoxes! Maybe we should be strict finitists.[8] We certainly should not reject otherwise compelling principles just because they break down in infinitary cases. I discuss this debate only to set it aside. The objection to regularity that I'll ultimately raise does not crucially depend on any reasoning about infinity. But to get there, we need to take a detour via a more general objection not to regularity, but to representing degrees of belief using probabilities in the first place. Ultimately, my response to that objection will invoke a picture of probabilistic reasoning that thoroughly rejects regularity.

3. Complexity

An influential objection to the Bayesian, probabilistic framework for representing belief and learning is that it is psychologically unrealistic because updating even a modestly complex body of belief by Bayesian conditionalization is computationally intractable. This line of thought is often associated with Harman (1986):

> I assume that, as far as the principles of revision we follow are concerned, belief is an all-or-nothing matter. I assume that this is so because it is too complicated for mere finite beings to make extensive use of probabilities (p. 27).

Harman supports this assumption with an argument to the effect that the Bayesian conditionalizer faces a problem of combinatorial explosion; as her corpus of beliefs gets bigger, the task of updating it by conditionalization gets much harder, very quickly. While Harman's argument for this claim was relatively informal, a few years later, computer scientist George Cooper (1990) proved that exact Bayesian inference is NP-hard, arguably vindicating Harman's suggestion.

Within philosophy, friends of probabilistic frameworks for thinking about belief and learning have tended to respond to considerations about the intractability of Bayesian inference by talking about ideals and idealization. Even if learning by conditionalization is not feasible for limited agents like us, it still represents a kind of rational ideal, and one that is useful for us to the extent that we can understand our own cognition as better or worse approximating it.[9] As I'll ultimately offer a very different response to worries about the intractability of probabilistic reasoning, I first want to say a bit about why I do not find this response satisfying.

There's an old joke that runs as follows. A tourist lost in a village in Scotland asks a local how to get to Aberdeen. The local responds: "Well, I wouldn't start from here." I suggest that *if* Harman is right that belief is an all-or-nothing matter for humans, then ideals concerning how to reason with degrees of belief are unhelpful in much the same way as the local's advice to the tourist is. It's no accident, perhaps, that David Christensen—one of the main philosophical defenders of the picture of Bayesian inference as a normative ideal—does *not* accept that humans do not have degrees of belief. Rather, as I understand him, he thinks we *do* have degrees of belief, and what's humanly intractable is ensuring that those degrees of belief are probabilistically coherent, and/or updated over time by conditionalization.[10] With this metaphysical picture of belief, we can imagine how a set

[7]For example, see Williamson (2007) and Pruss (2013). Hájek (2012) offers an interestingly different objection to regularity —namely, that it makes trouble for decision theory, since it leads to the result that all of our options must have infinite—or worse, undefined—expected utility. This is because there are coherent hypotheses in which we face gambles with infinite expected utility, and if regularity holds, we must assign those hypotheses some positive probability.

[8]For example, see Wright (1982).

[9]See, for example, Christensen (2004).

[10]This is how I interpret (Christensen, 2004, section 4.4).

https://doi.org/10.1017/can.2025.10015 Published online by Cambridge University Press

of degrees of belief that does not satisfy various coherence constraints might still get *closer* to satisfying them over time, and thus, how a picture of fully rational reasoning with degrees of belief could be a helpful kind of aspirational ideal.[11]

But here I want to offer a reply on Harman's behalf, and which I take to be congenial with what he wrote elsewhere.[12] In both economics and philosophy, the conceptual apparatus of subjective probabilities and utilities was introduced via representation theorems, which say, roughly, that *if* people's choices obey certain constraints, then they can be understood as having a probability function that represents their opinions, and acting so as to maximize the expectation (relative to their probability function) of a utility function that represents their values.[13] While this approach is not without opponents, and is probably more popular in economics than philosophy,[14] I think it's fair to refer to it as the orthodox approach to decision theory. And it requires that degrees of belief satisfy both synchronic and diachronic coherence constraints. Roughly, what this means is that *if* we want a broadly functionalist metaphysical picture of degrees of belief where representation theorems characterize the relevant functional roles, then we cannot easily accept that we have degrees of belief but they do not behave (even approximately) the way representation theorems say they do.[15] And if Harman is right that it's utterly intractable to update a large corpus of belief (which we have) by conditionalization, we should worry that any degrees of belief we might have could not behave even approximately the way representation theorems say they do.

So if we think the lesson to draw from Harman is that we do not have mental states that come anywhere near satisfying the constraints they'd need to *be* degrees of belief, then we should also think that norms about how to ideally reason with degrees of belief just aren't that relevant for us. Developing the skeptical line of thought, it's hard to see why a framework that describes an ideal of reasoning with degrees of belief should have any lessons concerning how to evaluate our *own* reasoning, which uses very different basic materials. To use a perhaps strained metaphor, knowing what ideal flight with flexible, flapping bird wings looks like might tell you almost nothing about how to design an airplane with rigid, immobile wings.

The response to Harman I'll consider is in important ways concessive; I want to grant to Harman the idea that norms for how an ideally rational agent would update an intractably vast corpus of degrees of belief are so far removed from human feasibility as to be uninteresting. But we can remain interested in norms of belief updating and decision-making that make extensive use of probabilities without accepting the picture Harman attacks. There is a tight connection between Harman's target—the picture of Bayesian norms describing how an ideal reasoner updates a vast corpus of belief—and regularity, because regularity forces the corpus of beliefs that an agent is updating to be vast. If it's coherent to suppose that I'm a brain in a vat, then regularity requires me to assign some small probability—perhaps infinitesimal—to this possibility. But quite a lot of hypotheses are coherent. If all coherent hypotheses must be assigned a positive probability, then by necessity the event space over which probabilities are distributed will be huge.[16] So, if we take Harmanian worries about computational tractability seriously, and try to respond to them by describing how probabilistic reasoning could be done using tractably small probability spaces, we are certain to end up describing a species of probabilistic reasoning that does not conform to a regularity principle.

[11]See, for example, Staffel (2019).

[12]For example, see Harman (1982).

[13]See, for example, Ramsey (1931), Von Neumann and Morgenstern (1947), Savage (1954)

[14]See Easwaran (2014).

[15]See, for example, Houtman and Maks (1985) and Varian (1991) for two ways of characterization how choices can approximate the constraints necessary for them to be representable as involving utility maximization. Both concern just utility, however, and not probability.

[16]This is a bit quick—I'll consider and reject a counterargument involving catch-all hypotheses in Section 3.

https://doi.org/10.1017/can.2025.10015 Published online by Cambridge University Press

Ultimately, that's where I'll go. I'll first try to show that everybody—not just the Bayesian—faces the problem of describing how updating beliefs can be tractable. In particular, I'll argue that the notorious "frame problem" is fruitfully viewed as a more general version of the problem of the complexity of updating probabilistic beliefs. In this more general setting, the problem is how an agent identifies a representation of her practical situation—one that contains enough relevant information for solving it to be fruitful while still being compact enough to be tractably analyzable—as the one to subject to some decision-making rule. This problem arises no less for an agent who reasons with all-or-nothing beliefs than for an agent who reasons with degrees of belief; reasoning with a large enough corpus of all-or-nothing beliefs is also intractably complicated. But if the problem can be solved for all-or-nothing belief, it can arguably be solved in a very similar way for degreed belief.[17] Ultimately, however, the solution will be incompatible with insisting that degrees of belief conform to a regularity principle.

While the original statement of the frame problem is due to McCarthy and Hayes (1981), I'll work with the broader formulation in Dennett (1990), who presents it in the form of a parable, which I'll quote in full:

> Once upon a time there was a robot, named R1 by its creators. Its only task was to fend for itself. One day its designers arranged for it to learn that its spare battery, its precious energy supply, was locked in a room with a time bomb set to go off soon. R1 located the room, and the key to the door, and formulated a plan to rescue its battery. There was a wagon in the room, and the battery was on the wagon, and R1 hypothesized that a certain action which it called PULLOUT (Wagon, Room, t) would result in the battery being removed from the room. Straightaway it acted, and did succeed in getting the battery out of the room before the bomb went off. Unfortunately, however, the bomb was also on the wagon. R1 knew that the bomb was on the wagon in the room, but did not realize that pulling the wagon would bring the bomb out along with the battery. Poor R1 had missed that obvious implication of its planned act.
>
> Back to the drawing board. 'The solution is obvious,' said the designers. 'Our next robot must be made to recognize not just the intended implications of its acts, but also the implications about their side-effects, by deducing these implications from the descriptions it uses in formulating its plans.' They called their next model, the robot-deducer, R1D1. They placed R1D1 in much the same predicament that R1 had succumbed to, and as it too hit upon the idea of PULLOUT (Wagon, Room, t) it began, as designed, to consider the implications of such a course of action. It had just finished deducing that pulling the wagon out of the room would not change the colour of the room's walls, and was embarking on a proof of the further implication that pulling the wagon out would cause its wheels to turn more revolutions than there were wheels on the wagon - when the bomb exploded.
>
> Back to the drawing board. 'We must teach it the difference between relevant implications and irrelevant implications,' said the designers, 'and teach it to ignore the irrelevant ones.' So they developed a method of tagging implications as either relevant or irrelevant to the project at hand, and installed the method in their next model, the robot-relevant-deducer, or R2D1 for short. When they subjected R2D1 to the test that had so unequivocally selected its ancestors for extinction, they were surprised to see it sitting, Hamlet-like, outside the room containing the ticking bomb, the native hue of its resolution sicklied o'er with the pale cast of thought, as Shakespeare (and more recently Fodor) has aptly put it. 'Do something!' they yelled at it. 'I

[17] I make a similar criticism not of Harman's argument against probabilistic frameworks for representing belief, but of a similar argument offered by Johnson et al. (2023), in Greco (2023a, 2023b).

https://doi.org/10.1017/can.2025.10015 Published online by Cambridge University Press

> am,' it retorted. 'I'm busily ignoring some thousands of implications I have determined to be irrelevant. Just as soon as I find an irrelevant implication, I put it on the list of those I must ignore, and…' the bomb went off (pp. 1–2).

While Dennett's discussion is rich and hard to summarize, one of the main points he makes is that positing that humans have a language of thought (LOT), a la Fodor (1975), does less work in explaining how we manage to intelligently cope with our environments than it might initially seem. R1, R1D1, and R2D1 all have languages of thought that they can use to represent propositions about their actions and the consequences of their actions. Still, they cannot intelligently act. While this is not meant to be an argument that an LOT is not a necessary piece of the puzzle, it is meant to show that in the absence of a solution to the frame problem—the absence of some story about how an agent manages to use their LOT to pose and analyze the right questions—we have not rendered our ability to flexibly respond to our environments in real time all that much less mysterious by positing an LOT.

I want to use the parable to make a similar point to Dennett's, but where the target is not the LOT hypothesis, but instead the hypothesis that human beliefs are all-or-nothing, rather than degreed. In Dennett's parable, each of R1, R1D1, and R2D1 is working with all-or-nothing and yes-no beliefs, rather than probabilities. Still, they either miss important implications of their beliefs or get bogged down in trivialities. It's not at all obvious what a solution to the frame problem would look like—or even whether it might be somehow circumvented rather than solved[18]—but a natural thought is that, *somehow*, intelligent agents manage to focus their cognitive resources on the right questions, such that answering those questions and acting on that basis is both computationally feasible, and practically fruitful. In effect, this natural thought amounts to positing a two-stage model of cognition in which the first stage is a black box that somehow solves the frame problem—that is, it identifies some set of questions as the relevant ones to answer—and the second stage analyzes the questions that the first stage has outputted, and then uses the answers to those questions as the basis for action. Once we conceive of things this way, I suggest that it's not clear why the second stage could not involve something that looks very much like textbook decision theory and expected utility maximization. If *everyone*—whether they think we have all-or-nothing beliefs, or degrees of belief—needs to appeal to a cognitive black box[19] that somehow identifies a tractable set of questions for analysis, it's not clear why those questions could not concern probabilities and expected utilities, rather than yes-no questions about the consequences of actions.

Before moving on, I want to preempt a natural concern. On a natural interpretation of Harman's argument, the problem is that probabilistic reasoning becomes very hard *very* quickly, so that even not-very-complex bodies of degreed belief are intractably hard to update by conditionalization. So, one might think that even if our frame-problem-solving black box gave us reasonably small problems of Bayesian inference to work with, they would *still* be intractable; so it must instead give us some yes-no questions to answer. This does not fit nicely with the lesson Cooper himself drew from his proof that computing exact posterior probabilities is NP-hard, however. Rather than concluding that probabilistic inference is hopeless, he concluded that "research should be directed away from the search for a general, efficient probabilistic inference algorithm, and toward the design of efficient special-case, average-case, and approximation algorithms" (Cooper, 1990, p. 393). And the subsequent literature in computer science has vindicated this suggestion—approximate Bayesian inference is now a rich field.[20] Roughly, the lesson, as I understand it, is that so long as you are willing to settle for methods for updating degrees of belief that give you output probabilities that are *close* to the true Bayesian posterior probabilities, then reasoning with

[18]Tyler Brooke Wilson (2025) considers and rejects the suggestion that modern machine learning methods show how the frame problem can be bypassed rather than solved, and I find his discussion persuasive.

[19]It does not *need* to be a black box. It's just that we currently do not have a great story about how it works.

[20]See Alquier (2020) for a survey.

https://doi.org/10.1017/can.2025.10015 Published online by Cambridge University Press

probabilities need not be computationally intractable. So, there's no obvious obstacle to the suggestion floated in the previous paragraph, as long as we are content to settle for the idea that the second stage of cognition is *approximately*, rather than perfectly, Bayesian.

In the next section, I'll flesh out the idea of a two-stage model of cognition, drawing on the distinction between small worlds and grand worlds, drawn by Savage (1954).

4. Small Worlds

Long before Harman or Cooper worried about the tractability of Bayesian updating, Savage (1954) grappled with very similar concerns in the context of decision theory. Savage endorses a policy of contingency planning, summarized by the proverb "look before you leap." But he immediately qualifies his endorsement:

> Carried to its logical extreme, the 'Look before you leap' principle demands that one envisage every conceivable policy for the government of his whole life (at least from now on) in its most minute details, in the light of the vast number of unknown states of the world, and decide here and now on one policy. This is utterly ridiculous, not—some might think—because there might later be cause for regret, if things did not turn out as had been anticipated, but because the task implied in making such a decision is not even remotely resembled by human possibility. It is even utterly beyond our power to plan a picnic or to play a game of chess in accordance with the principle, even when the world of states and the set of available acts to be envisaged are artificially reduced to the narrowest reasonable limits.
>
> Though the 'Look before you leap' principle is preposterous if carried to extremes, I would none the less argue that it is the proper subject of our further discussion, because to cross one's bridges when one comes to them means to attack relatively simple problems of decision by artificially confining attention to so small a world that the 'Look before you leap' principle can be applied there. I am unable to formulate criteria for selecting these small worlds and indeed believe that their selection may be a matter of judgment and experience about which it ia impossible to enunciate complete and sharply defined general principles (p. 16).

Savage does not appeal to considerations about combinatorial explosion or computational complexity to defend the necessity of confining one's attention to "small worlds," nor does he identify *probabilistic* reasoning in particular as the source of the intractability of grand-world reasoning. And he's right not to. To stick with one of his examples, grand-world reasoning about chess would not involve working with any probabilities. Ultimate consequences in chess—checkmates and stalemates—follow with certainty from ultimate board positions, and intermediate board positions follow with certainty from player decisions. But it's still intractable—for humans or computers—to play chess in the grand-world style, by considering the entirety of the game tree and using backward induction to select the optimal move at each point.

So, the lesson I take from Savage is that some kind of two-stage strategy for solving a decision problem—first, artificially confine one's attention to a tractable simplification of the problem and then solve it—is forced on us regardless of whether the decisions we face concern probability or not.

But—and we are now finally circling back to regularity—in the special case where the tractable problems we solve in the second stage *do* involve probability, I'll suggest that our probability assignments will inevitably violate a regularity principle, at least in spirit. Why should this be? One thought—too quick, but on the right track—is that the space of possibilities we consider in our small-world problem will have to rule out some coherent hypotheses. For instance, if we are deciding whether to call a bet in a game of poker, we are likely to think about a small world that includes each of the standard possibilities for what cards our opponents are holding, but which does *not* include possibilities in which they are holding nonstandard cards (e.g., an earl of quills, or a

https://doi.org/10.1017/can.2025.10015 Published online by Cambridge University Press

countess of clovers). Since it's coherent to imagine that someone is holding a card other than one of the standard 52, assigning such possibilities probability zero amounts to violating a regularity principle. One might object that there's a difference between assigning such possibilities probability zero, and simply ignoring them—leaving them out of the model. I think this objection rests on a misunderstanding. It's an axiom of probability theory that the state space, Ω, must be assigned probability 1. It's a quick consequence of the axioms that the complement of the state space, $0/$, must be assigned probability 0. In any given probability model, the event $0/$ can be interpreted as "none of the above." And the fact that it must be assigned probability 0 means that a poker model that "ignores" nonstandard possibilities does assign them probability 0, albeit as a class rather than case by case.

A natural suggestion at this point is that small-world probability assignments can avoid violating regularity by using a "catch-all" event—one which, by stipulation, occurs when none of the rest of the events considered in the small world occur.[21] As long as the catch-all is assigned some positive probability, then it would seem as if we have not violated regularity—if it turns out that your opponent was playing some elaborate practical joke which involved using nonstandard cards, that hypothesis wasn't ruled out—it was already accounted for, indirectly, in the catch-all.

There are familiar reasons to be unsatisfied with the use of catch-all hypotheses in the context of Bayesian confirmation theory. In general, it's very hard to motivate any specific assignments of probability to a catch-all hypothesis, let alone assignments of conditional probabilities of evidence given the catch-all; while we can offer considerations that bear on how likely we should be to make various astronomical observations given that Newtonian mechanics is true, or given that general relativity is true, it's much harder to say what the probabilities for observations should be given that some hypothesis other than any of the ones we have yet thought of is true. Similarly, in a decision theoretic context, it's very hard to motivate any assignments of payoffs to actions, given that the true state is the catch-all.

But I think we can sidestep these debates to argue that, whatever the merits of using catch-all hypotheses to leave room for hypotheses we have not imagined, they still do not really amount to a way of saving the spirit of the regularity principle. Imagine a small-world decision problem in which an agent A is considering two options, a_1 and a_2, with different payoffs depending on two states of the world, S_1 and S_2, along with a catch-all hypothesis C, for which she reserves 1% probability. This situation is represented below, with the question marks representing the difficulty of saying how fruitful various actions should be, conditional on some unimagined possibility turning out to be true.

	S_1 (0.49.5)	S_2 (0.49.5)	C (0.01)
a^1	1	0	?
a^2	0	1	?

While there are obvious difficulties with using such a small-world problem as a guide to action—how are we supposed to calculate expected utilities and decide what to do, given the unknowns?—I think we can also argue that the problem does not really respect the spirit of regularity. The decision table represents it as certain that if a_1 is chosen and the state of the world is S_1, the agent will receive a payoff of 1. But it's hard to see how we could interpret a_1 and S_1 such that the possibility of a different payoff, given that act and that state, is incoherent or impossible. Certainly, if the states are familiar states like, for example, what cards the opponent is holding, and the actions are familiar actions like calling a bet, it will be coherent to imagine that even though your opponent is holding worse cards than you, calling his bet will result in a loss (e.g., maybe he'll get so angry that he shoots you).

It seems to me the only way to avoid this problem is to pack enough detail into our states and actions that they really do jointly fix outcomes (which will also have to be impossibly detailed) with

[21]For a helpful, recent survey on the problem of new theories, in which such "catch-all" hypotheses are considered as a response to a problem for Bayesian epistemology, see Aronowitz (2025).

https://doi.org/10.1017/can.2025.10015 Published online by Cambridge University Press

necessity. But once we do this, states will be so specific and fine-grained that it will no longer be appropriate to call what we are working with a "small" world; any decision problem where states are individuated sufficiently finely to avoid the worry in the previous paragraph will have so much detail as to be intractable to solve.[22] So to insist for the sake of respecting regularity that actions, outcomes, and states should be specified with enough detail that we are *not* ignoring any possibilities by treating actions and states as together necessitating outcomes is itself to fail to take adequate account of our cognitive limitations.

Moreover, I want to suggest that once we accept what I've called the two-stage sketch of decision-making—that decision-making involves first somehow settling on a small-world problem, and then solving it—the motivation for thinking that small-world problems themselves should assign probability to a catch-all hypothesis is substantially diminished. Assigning some probability to a catch-all is meant to be a way of guaranteeing that we do not wrongly rule anything out. But a more attractive response to the risk of wrongly ruling things out, I suggest, is openness to revising which small-world model we are using, rather than insisting that all belief updating should be broadly like Bayesian conditionalization—updating within a model that's *already* made some allowance for the situation we find ourselves in. To insist that all updating should be modeled by a process like conditionalization—perhaps somehow involving a catch-all hypothesis—is to fail to respect our cognitive limitations. An agent is in a position to update by conditionalization when she's already anticipated some piece of evidence and has formed a contingency plan for how to respond to it. An agent who is limited in what evidence she's able to anticipate thus will not always be able to update by conditionalization. Because we cannot anticipate all possibilities, sometimes, rather than conditionalizing on evidence we have already made room for in our small-world model, we need to go back to the drawing board and come up with a new small-world model that can adequately account for the force of the previously unanticipated evidence.

This suggests me a more promising place to locate the kind of the open-mindedness characteristic of human intellectual humility that we have seen cannot be captured by a formal regularity principle. An intellectually humble human will not be too wedded to the small-world framing assumptions she uses to tackle a decision problem. That is, she'll have the right stance toward the fact that she is cognitively limited in what possibilities and evidence she can imagine. I admit that this is frustratingly vague as an account of the virtue of open-mindedness. But that vagueness is an unavoidable consequence of the fact that we do not, as yet, have an adequate account of what I've called the first stage of decision-making; our various formal decision theoretic methods tell us how to solve a given small-world decision problem, but not how to decide which small-world problem to solve, or when to revise such decisions. If the kind of the open-mindedness characteristic of intellectual humility manifests itself in this hard-to-formalize stage of inquiry, rather than in the stages that are better modeled using extant formal tools, it's no surprise that we lack a crisp characterization of it.

An optimist might think there's still an attractive, at least semiformal characterization of intellectual humility that survives the concession that obeying regularity is computationally intractable. While it's not the case that a rational agent never assigns coherent hypotheses probability zero, it *is* the case that for any coherent hypothesis a rational agent assigns probability zero, and there's *some* circumstance she could find herself in that would lead her to reconsider and start assigning it positive probability.

This thought strikes me as a potentially promising line of investigation, but no more than that. In the absence of a story about how rational agents do or should construct small-world decision

[22]Savage (1954) was well aware of the difficulties here:

…what are often thought of as consequences (that is, sure experiences of the deciding person) in isolated decision situations typically are in reality highly uncertain. Indeed, in the final analysis, a consequence is an idealization that can perhaps never be well approximated. I therefore suggest that we must expect acts with actually uncertain consequences to play the role of sure consequences in typical isolated decision situations (p. 84).

https://doi.org/10.1017/can.2025.10015 Published online by Cambridge University Press

problems—it's not even clear what we should characterize as the inputs to the process—it seems premature to insist on constraints like the one just mentioned. Maybe if we had a clear picture of how we *do* solve the frame problem, we'd also see that some hypotheses are "unreachable" for us[23]—we'd never start assigning them positive probability—but we'd also regard that as acceptable, because a different approach to solving that frame problem that would let us start assigning positive probability to those hypotheses would be, on the whole, worse for us. I have no idea.

5. Conclusion

I've argued that the open-mindedness central to intellectual humility cannot plausibly be captured by a regularity constraint on credences, given our cognitive limitations. Rather, open-mindedness—understood as the right stance toward our own fallibility—likely manifests in the way we approach the first-stage task of constructing tractable small-world models. But this task is not, as yet, remotely well understood.

Future work might take up the challenge of modeling or empirically investigating this first stage. What cues lead us to include or exclude outcomes, evidence, and world states in a small-world model? What prompts us to reconsider those inclusions and exclusions? If we can characterize the heuristics or patterns that guide these judgments, and in particular when they go well, we might be able to improve on pretheoretical, informal characterizations of open-mindedness.

Acknowledgments. Thanks to Brian Hedden, John Morrison, and Jonathan Vogel, as well as audiences at the Humility and Arrogance Conference as well as the University of Notre Dame for helpful feedback and discussion.

Daniel Greco is a professor at Yale University. While his interests are broad, the majority of his published research focuses on epistemology.

References

Alquier, P. (2020). Approximate bayesian inference. *Entropy*. 22(11), 1272

Appiah, K. A. (2017). *As if: Idealization and ideals*. Harvard University Press.

Aronowitz, S. (2025). The problem of new theories. In K. Sylvan, E. Sosa, J. Dancy, & M. Steup (Eds.), *The Blackwell Companion to Epistemology* (3rd ed). Wiley Blackwell.

Carr, J. (2021). *Why ideal epistemology?* Mind.

Carroll, S. (2017) Bayes's theorem. *2017: What scientific term or concept ought to be more widely known?* . Edge , 2017. https://www.edge.org/response-detail/27098.

Christensen, D. (2004). *Putting logic in its place*. Oxford University Press.

Christensen, D. (2007). Does Murphy's law apply in epistemology? Self-doubt and rational ideals. In T. S. Gendler & J. Hawthorne (Eds.), *Oxford Studies in Epistemology* (Vol. I, pp. 3–31). Oxford University Press.

Cooper, G. F. (1990). The computational complexity of probabilistic inference using bayesian belief networks. *Artificial Intelligence*. 42(2-3), 393–405

Dennett, D. C. (1990). Cognitive wheels: The frame problem of AI. *The Philosophy of Artificial Intelligence*, 147, 1–16.

Easwaran, K.. (2014). Decision theory without representation theorems. *Philosophers' Imprint* 14, 1–30.

Easwaran, K., Hájek, A., Mancosu, P., & Oppy, G. 2021. Infinity. *Stanford Enyclopedia of Philosophy*. Co Principal Editors: Edward N. Zalta and Uri Nodelman The Metaphysics Research Lab Philosophy Department Stanford University Stanford, CA 94305–4115 World Wide Web URL: https://plato.stanford.edu/Publisher:

Fodor, J. (1975). *The language of thought*. Harvard University Press.

Greco, D. (2023a). The small world's problem is everyone's problem, not a reason to favor cnt over probabilistic decision theory. *Behavioral and Brain Sciences*, 46, e95. doi:10.1017/S0140525X22002771>.

Greco, D. (2023b). *Idealization in epistemology: A modest modeling approach*. Oxford University Press.

Hájek, A. (2012). Is strict coherence coherent? *Dialectica*, 66(3), 411–424. http://doi.org/10.1111/j.1746-8361.2012.01310.x.

Harman, G. (1982). Conceptual role semantics. *Notre Dame Journal of Formal Logic*, 23(2), 242–256.

Harman, G. (1986). *Change in view: Principles of reasoning*. The MIT Press.

[23]Perhaps akin to the "blind spots" discussed by Sorensen (1988).

https://doi.org/10.1017/can.2025.10015 Published online by Cambridge University Press

Houtman, M., & Maks, J. (1985). Determining all maximal data subsets consistent with revealed preference. *Kwantitatieve Methoden*, 19(1), 89–104.
Jackson, E. G. (2020). The relationship between belief and credence. *Philosophy Compass*, 15(6), e12668.
Johnson, S. G. B., Bilovich, A., & Tuckett, D. (2023). Conviction narrative theory: A theory of choice under radical uncertainty. *Behavioral and Brain Sciences*, 46, e82.
Leitgeb, H. (2014). The stability theory of belief. *Philosophical Review*, 123(2), 131–171.
Lewis, D. (1980). A subjectivist's guide to objective chance. In R. C. Jeffrey (Ed.), *Studies in Inductive Logic and Probability* (Vol II). University of California Press.
McCarthy, J., & Hayes, P. J. 1981. Some philosophical problems from the standpoint of artificial intelligence. In Morgan, Kaufmann (ed.), *Readings in artificial intelligence* (pp. 431–450). Elsevier.
Moon, A. (2017). Beliefs do not come in degrees. *Canadian Journal of Philosophy*, 47(6), 760–778.
Moss, S. (2013). Epistemology formalized. *Philosophical Review*, 122(1), 1–43.
Pruss, A. R. (2013). Probability, regularity, and cardinality. *Philosophy of Science*, 80(2), 231–240.
Ramsey, F. 1931. Truth and probability. In *The foundations of mathematics and other logical essays* (pp. 122–130). Harcourt, Brace, and Co.
Savage, L. J. (1954). *The foundations of statistics*. Wiley Publications in Statistics.
Snow, N. E. 2018. Intellectual humility. In Heather, Battaly (ed.), *The Routledge handbook of virtue epistemology* (pp. 178–195). Routledge.
Sorensen, R. *Blindspots*. 1988 Oxford University Press.
Staffel, J. (2019). *Unsettled thoughts: A theory of degrees of rationality*. Oxford University Press.
Sturgeon, S. (2020). *The rational mind*. Oxford University Press.
Thorstad, D. (2024). *Inquiry under bounds*. Oxford University Press.
Varian, H. R. (1991). *Goodness-of-fit for revealed preference tests*. Department of Economics, University of Michigan Ann Arbor.
Von Neumann, J., & Morgenstern, O. 1947. *Theory of games and economic behavior* (2nd rev.) Princeton University Press.
Whitcomb, D., Battaly, H., Baehr, J., & Howard-Snyder, D. (2017). Intellectual humility: Owning our limitations. *Philosophy and Phenomenological Research*, 94(3), 509–539. http://doi.org/10.1111/phpr.12228.
Williamson, T. How probable is an infinite sequence of heads? *Analysis*, 67(3):173–180, 2007. http://doi.org/10.1093/analys/67.3.173. http://analysis.oxfordjournals.org.
Wilson, T. B. (2025). *Foundations for a science of central cognition* [Unpublished manuscript].
Wright, C. (1982). Strict finitism. *Synthese*, 51(2), 203–282. http://doi.org/10.1007/bf00413828.
Yap, A. (2014). Idealization, epistemic logic, and epistemology. *Synthese*, 191(14), 3351–3366.
Zhang, S. (2024). Approximate rationality and ideal rationality. *Asian Journal of Philosophy*, 3(2), 45.

Cite this article: Greco, D. 2025. Humility and Complexity. *Canadian Journal of Philosophy*: 1–12, doi:10.1017/can.2025.10015