Non-Ideal Epistemic Rationality[*]

Nick Hughes
University of Oslo
nickhowellhughes@gmail.com

ABSTRACT: I develop a broadly reliabilist theory of non-ideal epistemic rationality and argue that if it is correct we should reject the recently popular idea that the standards of non-ideal epistemic rationality are mere social conventions.

§1. IDEAL AND NON-IDEAL EPISTEMOLOGY

Here's something that won't come as news: some people have greater abilities than others. Turner could paint light beautifully, Maria Callas could sing a high C note, and Simone Biles can do a double backflip. Most of us can't do any of these things.

As it is with painting, singing, and gymnastics, so it is with thinking. Here are two examples:

> **THEOREM X**: Let 'X' be an as-yet-undiscovered and extremely hard to prove mathematical theorem. Assume that it's nevertheless knowable *a priori*. Since X is knowable *a priori*, a cognitively ideal agent – call her 'SuperMind' – will believe that it's true, having effortlessly produced a proof of it. But I can't do that – it's far beyond my abilities. So instead, I suspend judgement on it. And I take myself to be rational in doing so.

[*] Forthcoming in *Philosophical Issues*.

> **RAIN IN PARIS:** I've acquired a large and complex body of evidence bearing on whether it will rain in Paris next Wednesday. Call the proposition that it will rain 'p'. The evidence, let's suppose, makes it exactly 0.67745 likely that p. A cognitively ideal agent like SuperMind will believe that p is exactly 0.67745 likely. But I can't figure out the probability to anything like that level of precision. Instead, I only believe that it's between 0.5 and 0.8 likely. Again, I take it that I'm rational in doing so.

Ideal Epistemology is a research programme that idealizes away from cognitive limitations and studies norms of rational thinking for cognitively ideal, or close to ideal, agents. That is, the kinds of agents, like SuperMind, who would effortlessly produce a proof of Theorem X and would have a belief about the likelihood of it raining in Paris next Wednesday exactly matching the probability on their evidence.

Cognitively ideal agents are typically assumed to have:[1]

- Unlimited computational power.
- Unlimited informational storage space.
- The ability to process information instantaneously.
- Perfect retention and recall.
- A complete absence of biases.
- …and various other extraordinary abilities and capacities.

Here are some of the claims that have been made about them in the literature:

- They are logically omniscient.
- They believe all *a priori* truths.
- Their beliefs perfectly conform to their evidence.
- Their credences are probabilistically coherent.
- They have infinitely precise credences.
- They update by conditionalization.

[1] I've adapted the lists below from Carr (2022)

- They update their beliefs instantaneously.

But of course, none of us is, or ever will be, cognitively ideal. The rap sheet is long. Here are some of the lowlights:

- We have limited computational power.
- We have limited storage space.
- We have limited processing speeds.
- We have imperfect retention and recall.
- We're prone to fallacious reasoning.
- We have all kinds of biases.
- We jump to conclusions.
- …and so on.

And as a result:

- We're not logically omniscient.
- We don't believe all *a priori* truths.
- Our beliefs don't perfectly conform to our evidence.
- Our credences aren't probabilistically coherent.
- Our credences are not infinitely precise.
- We usually don't update by conditionalization.
- We often update slowly or even not at all.

This gives rise to a question: given that we can't think like cognitively ideal agents, how should we think? Answering this question is the job of a theory of non-ideal (aka bounded) epistemic rationality.[2] In this paper, I'll sketch the rough outlines of an answer to the question that I think is quite plausible.

Existing approaches to non-ideal rationality mostly fall into roughly two camps. In the first camp there are non-ideal Bayesians. They try to explain non-ideal rationality by positing norms that, in one way or another, relax or approximate

[2] Whenever I talk about "non-ideal rationality", I'll mean non-ideal *epistemic* rationality.

those of standard Bayesianism.[3] In the second camp are advocates of 'ecological rationality' like Gerd Gigerenzer.[4] Though they usually focus on practical, rather than epistemic rationality, they sometimes argue that in certain environments it can be rational to form beliefs using fast and frugal heuristics, even though the resulting beliefs systematically violate Bayesian norms.[5]

The kind of approach I'll be developing here doesn't obviously fit into either of these camps. Instead, I'll be building on some ideas from the reliabilist tradition in epistemology.[6]

One thing I want to flag immediately, however, is that, as will soon become apparent, I'm using "reliabilism" in quite a broad way. Any theory of epistemic rationality that evaluates beliefs as rational or irrational by looking at how the way they are formed and held fares at getting it right about p (e.g., having a true belief about p or a knowing that p) across a range of possible worlds will count as "reliabilist" on the way I'm using it.[7] This means that even some views that are sometimes thought of as being competitors to reliabilism, like dispositionalism,[8] various kinds of knowledge-first epistemology,[9] and various kinds of virtue epistemology,[10] will count as versions of reliabilism.

It may be that the reliabilist approach ends up dovetailing quite nicely with some ideas from the ecological rationality camp. As Rysiew (2008) and Gigerenzer (2021)

[3] E.g., Hacking (1967), Garber (1983), Staffel (2019), Skipper & Bjerring (2020), Pettigrew (2021).

[4] E.g., Gigerenzer *et al* (1999), Gigerenzer & Selten (2002).

[5] For other work on non-ideal epistemic rationality, see Cherniak (1990), Morton (2012), Smithies (2015, 2022), Ballantyne (2019), DiPaolo (2019), Gao (2019, 2023), Begby (2021), Dorst (2020, 2023), Singer (2023), Thorstad (fc, ms), Hirvela (ms), Icard (ms), and Lasonen-Aarnio (ms).

[6] My approach shares some affinities with Smithies (2015, 2022) and Lasonen-Aarnio (ms), but also differs from these accounts in many respects.

[7] I'm using "getting it right" as a semi-technical term here. One gets it right about p iff one has the objectively epistemically correct doxastic attitude towards p. Traditionally it has been thought that a belief is objectively correct iff it is true. But, for reasons I'll explain below, I want to leave it open that a belief is objectively correct iff one knows that p.

[8] Lasonen-Aarnio (2010, 2014, 2020, 2021, fc1, fc2, ms), Williamson (fc1), Hughes (2023)

[9] E.g., Williamson (fc1) and Simion (2024a, 2024b).

[10] E.g., Beddor & Pavese (2020)

have pointed out, a lot of work on ecological rationality is reliabilist in spirit.[11] It may even be possible to square it with a Bayesian approach (though this is less clear). But I won't go into any of this here.

§2. THE PLAN

The plan is this.

In the first half of the paper, I'll sketch the outlines of a reliabilist theory of non-ideal rationality and show how it does a nice job of explaining the data in the THEOREM X and RAIN IN PARIS cases, which is at least some evidence that the framework is on the right track.[12]

Then, in the second half of the paper, I'll discuss the recently popular idea that the standards of non-ideal rationality are mere social conventions. I'll argue that my theory, which denies this, does a better job of explaining the data that has been cited in favour of conventionalism.

One thing to note. A lot of work in non-ideal epistemology, especially within the Bayesian tradition, focuses on fine-grained attitudes like credences. But I'm only going to discuss the coarse-grained attitudes (that is, belief, suspension, and disbelief). I hope it will be possible to extend the framework to cover credences. But I won't try to do that here.

§3. A RELIABILIST FRAMEWORK

[11] As Rysiew (2008) and Gigerenzer (2021) have pointed out, a lot of work on ecological rationality is reliabilist in spirit. For a recent proposal on how to develop reliabilism in a way that enables it to ground the ecological rationality programme, see Dusi (fc).

[12] For the most part, I won't respond to objections to general objections to reliabilism that aren't specific to my implementation of it here (e.g. Cohen's (1984) New Evil Demon argument). I think they have been adequately addressed by others.

Here's the data that needs to be explained in the THEOREM X and RAIN IN PARIS cases:

- It's rational for SuperMind to believe that Theorem X is true. But it's not rational for me to believe this. Rather, I should suspend judgment on it.[13]

- It's rational for SuperMind to believe that rain is exactly 0.67745 likely. But it's not rational for me to believe this. Rather, I should only believe that it's between 0.5-0.8 likely.

Why does rationality pattern in this way? I think that the answer has something to do with the differences in our ability to reliably get it right about these things. If I were to form a belief about Theorem X or the precise probability of rain, I'd just be guessing. Given my cognitive limitations, I can't reliably get these things right. But SuperMind wouldn't just be guessing. She doesn't have any cognitive limitations, so she can reliably get them right. And it's these facts about our differing reliability, I suggest, that explain the data.

That's the short answer. Now for the longer answer. I haven't figured out all the details yet by any means, but here's a rough general framework for thinking about non-ideal rationality within a reliabilist ideology. It basically says that if you can have a reliably held belief about whether p, you should, and if you can't, you shouldn't. Call this the "core idea":

> **CORE IDEA**: It is rational to believe that p iff you can have a reliably held belief about whether p.

The framework has four conditions. Two on *ex post* rationality, and two on *ex ante* rationality. As is usual with reliabilist theories, *ex post* rationality is in the driving seat. So, first:

[13] I will assume that one ought to be rational, and so that we can move freely between talk of what it's rational for you to do, what you ought to do, and what you're required to do.

EX POST RATIONALITY

> **RB**$^{\text{EP}}$: S’s belief about p is *ex post* rational iff the way that she comes to hold the belief reliably results in her getting it right about p. Otherwise, her belief is irrational.[14,15]

> **RS**$^{\text{EP:}}$ S’s suspending judgement on p is *ex post* rational iff she suspends on p for the reason that she can’t come to have a belief about p in a way that reliably results in her getting it right about p. Otherwise, her suspending judgement is irrational.[16]

More succinctly: your belief is rational iff it is reliably held, and your suspending judgement is rational iff you suspend because you can’t have a rational belief.

Given the priority of *ex post* rationality, it is natural to think of *ex ante* rationality as a matter of it being possible for you to have an *ex post* rational attitude. In other words, you’re permitted to take doxastic attitude D to p iff you can rationally take D to p. Hence:

EX ANTE RATIONALITY

> **RB**$^{\text{EA}}$: It’s *ex ante* rational for S to have a belief about p iff she can have a belief about p that’s held in a way that reliably results in her getting it right about p. Otherwise, it’s *ex ante* irrational for S to have a belief about p.[17]

[14] This claim isn’t new, of course. It’s the core thesis of bread-and-butter reliabilism, going back to Goldman (1979).

[15] Hereafter I’ll use “reliably held belief”, “reliably formed belief”, etc., as shorthand for “held in a way that reliably results in one getting it right”.

[16] As will become clear soon, “reason” is being used here in a causal-explanatory sense, not in a normative sense.

[17] Note the similarity to Goldman’s (1979) proposal that: “Person S is *ex ante* justified in believing p at t if and only if there is a reliable belief-forming operation available to S which is such that if S applied that operation to his total cognitive state at t, S would believe p at t-plus-delta (for a suitably small delta) and that belief would be *ex post* justified”

> **RS^{EA}**: It's *ex ante* rational for S to suspend judgement on p iff she can suspend judgement on p for the reason that she can't come to have a belief about p that's held in a way that reliably results in her getting it right about p. Otherwise, it's *ex ante* irrational for S to suspend judgement on p.

More succinctly: you may believe that p iff you can have a reliably held believe that p, and you may suspend judgement on p iff you can suspend on p because you can't have a reliably held belief.

So, that's the framework. In §5 I'll argue that it makes the right predictions about the THEOREM X and RAIN IN PARIS cases. In short: it predicts that SuperMind is rationally required to believe that Theorem X is true, and that rain is exactly 0.67745 likely, but that I'm rationally required to suspend judgement on these things.

But before I do that, I want to clarify some things about the framework.

§4. CLARIFYING THE FRAMEWORK

§4.1. HOW SHOULD WE FILL IN THE DETAILS?

First, for the most part I'm going to leave the details of the interpretation of the framework very open here. In particular, I'm not going to take a stand on how we should think about what it is for a way of holding a belief to be *reliable*, what it is to *get it right*, or on what *ways* are and how they should be individuated.

Let me explain, starting with reliability. The framework talks about ways of coming to have beliefs being reliable or not. But of course, there are many different things one might have in mind when one talks about reliability. One question is: in what *circumstances* must a way of coming to have a belief get it right in order to count as reliable? So, for instance, we might think that, in order to count as a reliable, a way of coming to have a belief must get it right about p in the actual world,[18] normal worlds,[19] normal circumstances,[20] non-manipulated circumstances,[21] or some other set of worlds or circumstances. And even within these categories, different views are possible. For example, Goldman (1986) takes normal worlds to be worlds that share the general characteristics that we take the actual world to have, whereas Beddor & Pavese (2020) take normal worlds for the performance of a task to be worlds where it would be fair to positively or negatively evaluate the performance. I won't take a stand on these issues here.

Independently of what worlds or circumstances one thinks a way of coming to have a belief must get it right in, there is also the question of *how often* the way of coming to believe must get it right to count as reliable. Again, there are different possible views here. For instance, Goldman (1979) takes a way of coming to have a belief to be reliable iff it yields a sufficiently high ratio of true to false beliefs above some threshold < 1. But others have argued that it is better to think of a way of coming to have a belief as being reliable iff it *never* yields false beliefs (e.g., Dretske 1981). Again, I won't take a stand on this issue here.

Nor are we forced to think of "getting it right" to be a matter of having a true belief rather than a false belief, as is standard in reliabilist theories. According to some knowledge-first theorists, one gets it right in believing that p iff one knows that p. [22] True belief alone isn't enough. Once again, I won't take a stand on this here.

What are "ways" of coming to have a belief? Again, different views are possible. Goldman (1979) takes them to be processes. Lyons (2019) also takes them to be processes and proposes an "algorithms-and-parameters" scheme for individuating them. Lasonen-Aarnio (2021, fc1) argues that they should be identified with the

[18] Goldman (1979)

[19] Goldman (1986)

[20] Graham (2017)

[21] Goldman (1979)

[22] Williamson (2000, fc1), Littlejohn (2013, fc), Sutton (2007).

cognitive dispositions the agent manifests in coming to believe that p. Other views are also possible. Again, I won't take a stand.[23,24]

As I mentioned earlier, an upshot of all this is that some theories that are not usually described as "reliabilist" will come out as possible interpretations of the framework. For example: dispositionalism, some knowledge-first theories, and some virtue-theoretic approaches.

I'm leaving the framework schematic here for two reasons.

Firstly, because my aim here isn't to try to figure out and defend the most plausible version of reliabilism (for one thing, that's far too large a task for a single paper). Rather, it is to show how the reliabilist way of thinking can furnish us with a plausible theory of non-ideal rationality.

Secondly because, for the most part, nothing I want to say turns on spelling out the details one way rather than another.[25] Many different versions of reliabilism are able to capture the observation that there's no way for me – but is a way for SuperMind – to come to have reliably held beliefs about Theorem X or the precise probability of rain. It doesn't matter, for example, whether we think the kind of reliability that matters involves getting it right in the actual world, in normal worlds, or in non-manipulated worlds, or whether we think reliability requires never getting it wrong, or just mostly getting it right, or whether we think of getting it right as a matter of having a true belief or knowledge. So, I don't need to take a stand on the details.[26]

[23] Since Theorem X is necessarily true, however, they will have to be thought of in a somewhat coarse-grained way in order to capture the datum that there is no way for me to come to have a reliably held belief about the truth-value of the Theorem.

[24] A common objection to reliabilism is that, because there is no principled way to individuate ways of forming beliefs, and whether a belief is reliably formed or not depends on the way that it was formed, it's impossible to say whether a given belief is reliably formed or not. This is the generality problem. I am not moved by this objection, for the same reasons as Goldman (2021): we don't actually need a principled way of individuating ways of forming beliefs in order to tell whether a given belief was reliably formed or not.

[25] There will be a couple of exceptions. I'll discuss these later.

[26] For what it's worth, I prefer a Lasonen-Aarnio-style knowledge-first dispositionalist approach to reliabilism (Hughes 2023, 2024). In Hughes (2023) I argue that this approach can shed light on what's going

§4.2. RATIONALLY SUSPENDING

Secondly, I need to clarify RS$^{\text{EP}}$ which says that:

> **RS$^{\text{EP}}$**: S's suspending judgement on p is *ex post* rational iff she suspends on p for the reason that she can't come to have a belief about p in a way that reliably results in her getting it right about p. Otherwise, her suspending judgement is irrational.

RS$^{\text{EP}}$ is motivated by the thought that rational agents don't suspend judgement willy-nilly. Rather, they suspend because having a belief is epistemically out of bounds. RS$^{\text{EP}}$ implements that idea on the assumption that believing is in bounds iff one can have a belief about p that's held in a way that reliably results in one getting it right about p

At first glance, RS$^{\text{EP}}$ might seem to overintellectualize rational suspension. It's not like you have to go through the process of thinking to yourself "I can't reliably get it right about whether p, so I'll suspend judgement" in order to rationally suspend on something. But RS$^{\text{EP}}$ could be read as demanding exactly that.

That's not the interpretation of RS$^{\text{EP}}$ that I have in mind. All that's needed, as I'm thinking of it, is that when you suspend judgement on p you exhibit an appropriate *sensitivity* to your ability to have a reliably held belief. As a rough guide, we can say that you rationally suspend on p iff: if you could have had a reliably formed belief about p, then you would have done.[27]

§4.3. WHAT CAN YOU DO?

The third thing I want to clarify is this. At first glance, the framework might seem wildly demanding. Consider RS$^{\text{EA}}$ again:

wrong in epistemic feedback loops – situations where a prior attitude towards p causes you to manufacture evidence in favour of p, which then causes you to believe that p.

[27] Though this cannot be exactly right. As an analysis it would commit the conditional fallacy.

> **RS^{EA}**: It's *ex ante* rational for S to suspend judgement on p iff she can suspend judgement on p for the reason that she can't come to have a belief about p that's held in a way that reliably results in her getting it right about p. Otherwise, it's *ex ante* irrational for S to suspend judgement on p.

A consequence of this is that if you can have a reliably held belief about whether p, then you're rationally required to believe that p rather than suspend judgement. And that might seem totally implausible.

§4.3.1. The Context-Sensitivity of "Can"

The reason has to do with the context-sensitivity of the "can" in "you can have a reliably held belief about whether p". On the standard Lewis-Kratzer semantics for agentive modals,[28] "S can φ" is true iff S's φ-ing is compossible with the relevant set of facts being held fixed. But what facts are held fixed varies across contexts. Lewis (1976) gives this example:

> "An ape can't speak a human language – say, Finnish – but I can. Facts about the anatomy and operation of the ape's larynx and nervous system are not compossible with his speaking Finnish. The corresponding facts about my larynx and nervous system are compossible with my speaking Finnish. But don't take me along to Helsinki as your interpreter: I can't speak Finnish. My speaking Finnish is compossible with the facts considered so far, but not with further facts about my lack of training. What I can do, relative to one set of facts, I cannot do, relative to another, more inclusive, set." (1976: 149)

"Lewis can speak Finnish" is true in a context where only the facts about his larynx and nervous system are held fixed, but false in a context where his lack of training in Finnish is also held fixed.

[28] Lewis (1976), Kratzer (1977).

I've claimed that I can't prove Theorem X. But in some contexts this claim is false. If I were to study mathematics for many years and take not-yet-existing cognition-enhancing drugs, then I could prove it. My claim that I can't prove it is only true holding fixed the fact that I only have high school maths and don't have access to cognition-enhancing drugs.[29]

The worry is that if we interpret the framework's claim that if you can have a reliably held belief about whether p, then you're rationally required to believe that p, rather than suspend judgement, in a way on which "can" is relatively unrestricted, so that little is held fixed, then the framework delivers the result that I am rationally required to believe that Theorem X is true, rather than suspend judgement on it. And that is clearly the wrong result – it makes the framework far too demanding.

§4.3.2. Privileged Worlds

There is a solution to this problem. For whilst there is an interpretation of "can" on which I can have a reliably held belief about whether Theorem X is true, this interpretation is not relevant to determining whether I'm required to believe it. Only the interpretation of it on which I can't have a reliably held belief is relevant.

Why? Because there is, I suggest, a *privileged* interpretation of "can" in non-ideal rationality. Consider the familiar idea of nearby and distant possible worlds. World A is nearby to world B to the extent that world A is similar to world B. World A is distance from world B to the extent that world A is dissimilar to world B. What I propose is that interpretations of "can" in "S can have a reliably held belief about whether p" that include worlds *distant* to the world the cognitively non-ideal agent inhabits are irrelevant to determining whether the agent is rationally required to believe that p. When it comes to rationality, all that matters is whether the agent has a reliably held belief in *nearby* worlds. Or, putting it another way, when it comes to rationality, interpretations of statements of the form

[29] For discussion of this point in relation to the concept of being-in-a-position-to-know, see Kearl & Willard-Kyle (fc)

“S can have a reliably held belief about whether p” on which we restrict ourselves to what happens nearby worlds are *privileged*. Call this “privileged worlds theory”:

> **PRIVILEGED WORLDS THEORY**: S is rationally required to believe that p only if S has a reliably held belief in worlds nearby to the world S inhabits.

In the THEOREM X case, worlds in which I study mathematics for many years and take not-yet-existing cognition-enhancing drugs are very dissimilar to, and hence distant from, the world I inhabit. All the similar, and so nearby, worlds are worlds in which I only have high school maths and don’t have access to cognition-enhancing drugs. And in all of these worlds, if I form a belief about the Theorem, my belief is nothing more than a guess, and so not reliably held. Hence, privileged worlds theory delivers the correct result that I am not rationally required to believe that Theorem X is true. So, by supplementing the framework with privileged worlds theory, we avoid the demandingness worry here.

§4.3.3. More demandingness worries

The demandingness worry just discussed concerns intellectually demanding propositions. Privileged worlds theory deals with it. But this isn’t the only worry.

Imagine that you’re in a taxi, and as the driver is trying to navigate a busy roundabout, you ask him “what does 9 x 12 equal?’. He says “I can’t think about that right now, I’m trying to concentrate on driving” and he suspends judgement. This is surely fine from a rational point of view. But there is an interpretation of “can” on which the taxi driver clearly can form a belief about the answer to the question in a way that reliably results in him getting it right – all he needs to do is use elementary arithmetic, which we can assume he possess. Moreover, this proposition is not intellectually demanding in the same way as Theorem X. So we don’t need to imagine him studying mathematics for many years and take not-yet-existing cognition-enhancing drugs to reach a world where he has a reliably held belief about the answer. Worlds where he has a reliably held belief are not very dissimilar to the world he inhabits; all he has to do is pull over and think about it for a few seconds. Even taking into account privileged worlds theory, then, it might seem that, according to the framework, he is rationally required to have a belief about the answer. But that is the wrong result; he's fine suspending judgement.

Cases like this can be multiplied. Right now, I suspend judgement on whether the population of Canada is over 40 million people. And I take it that I'm rational in doing so. But I could very easily find out – all I need to do is open my web browser and google it. Google is reliable about these things, so there's an interpretation of "can" on which I can form a belief about it in a way that reliably results in me getting it right. And again, worlds where I do so aren't very dissimilar to the world I inhabit. So it seems that, according to the framework, I'm rationally required to have a belief about the matter. Again, that is the wrong result.

These cases are instances of a general problem that besets a certain kind of normative epistemology. RS^{EA} means that the framework doesn't just posit negative epistemic requirements of the form 'you must: believe that p only if conditions C obtain', but also positive epistemic requirements of the form 'you must: believe that p, rather than suspend on p, if conditions C obtain'.[30] Theories that posit positive epistemic requirements run the risk of being overdemanding owing to what Mark Nelson (2010) calls the "infinite justificational fecundity of evidence". The worry is that, even putting intellectually demanding propositions aside, for any conditions C an epistemologist might posit (like, say, "you can have a reliably held belief about p", "you're in a position to know that p", or "your evidence indicates that p") those conditions will obtain for so many propositions that you will never be able to satisfy all, or even a fraction, of the resulting requirements. Nor does it seem, from the point of view of rationality, that you should have to; the taxi driver is doing fine epistemically in suspending on 9 x 12, and I'm doing fine suspending on what the population of Canada is. The challenge is to stop positive epistemic requirements from proliferating in this way. Call it the "containment challenge":

> **THE CONTAINMENT CHALLENGE:** Any epistemology that posits positive epistemic requirements must ensure that they don't proliferate wildly.

[30] For other proposals about positive epistemic norms, see Ichikawa (2022) and Simion (2024a, 2024b).

I think that my framework can meet the containment challenge. I suggest that the privileged interpretation of the "can" statements in these norms goes beyond privileged worlds theory. It isn't *all* nearby worlds that matter when it comes to rationality, it's only a subset of them. The taxi driver isn't rationally required to have a belief about what 9 x 12 equals while he's driving on a busy roundabout because there's no nearby world where he has a reliably held belief *compatible with him concentrating on driving*. I'm not rationally required to have a belief about whether the population of Canada is over 40 million people because there's no nearby world where I have a reliably held belief *compatible with me not opening my browser and googling the answer*. Worlds where the driver doesn't concentrate on driving and worlds where I open my browser and google the answer are not part of the privileged subset.

This raises a question: what factors contribute to determining the privileged subset of nearby worlds relevant to rationality? Presumably they'll include things like whether you're busy doing other things and where your attention is focused. But it is unrealistic to expect that we will be able to first figure out what factors contribute (and when and to what extent) and then use our results to read off a verdict about what attitude it's rational for you to take towards p in a given case. Rather, we should approach the issue from the other direction, using our prior judgement about what attitude it's rational for you to take towards p to decide what factors contribute, when, and to what extent.[31]

One thing that's clear is that not just any old factor can be thrown in. Suppose that you have overwhelming evidence that your friend is guilty of murder, but you don't believe it because you don't want to. It's not plausible to say that you're rational in suspending judgement on whether your friend did it because you can't have a reliably held belief about whether he's guilty compatible with you believing what you want to believe.

[31] Compare with Williamson's (2000) suggestion that, on the safety theory of knowledge, we shouldn't try to first figure out what worlds count as nearby and use this to generate results about what you know. Instead, we should use our judgements about what you know to figure out what worlds count as nearby.

What do the legitimate restricting factors have in common then, that the illegitimate ones don't? That is, what explains why it's rational for the taxi driver to suspend judgement on 9 x 12 and rational for me to suspend judgement on the population of Canada, but not rational for you to suspend judgement on whether your friend is guilty of murder? I don't know. I think it would be very interesting to find out. In fact, it's a major task for the development of the framework. But it's a task for another time. All I want to point out here is that we can invoke those factors that *do* make a difference to meet the containment challenge.

§5. THE FRAMEWORK'S PREDICTIONS

With these clarifications in place, I'll now go through the predictions that the framework makes about the THEOREM X and RAIN IN PARIS cases. In each case, the prediction is, I submit, the right one.

§5.1. THEOREM X

As we've seen, I can't have a reliably held belief about Theorem X. Given my cognitive limitations, in all nearby worlds where I have a belief about it I am guessing.[32] But I can suspend on it for the reason that I can't have a reliably held belief. So, the framework says that:

- It's *ex ante* rational for me to suspend judgement on the Theorem and *ex ante* irrational for me to believe it (by RS^{EA} and RB^{EA}).

- If I suspend on it for the reason that I can't have a reliably held belief, then I suspend rationally (by RS^{EP}).

But SuperMind can have a reliably held belief about it. So, the framework says that:

[32] Hereafter, whenever I talk about someone being able, or not able, to have a reliably held belief, I will mean that there are, or are not, *nearby* worlds where they have one, unless I say otherwise.

- It's *ex ante* rational for SuperMind to believe that the Theorem is true, and *ex ante* irrational for her to suspend on it (by $\mathrm{RS}^{\mathrm{EA}}$ and $\mathrm{RB}^{\mathrm{EA}}$).

- If SuperMind forms her belief about the Theorem's truth-value in a reliable way, then her belief is rational (by $\mathrm{RB}^{\mathrm{EP}}$).

I think these are the right predictions.

§5.2. RAIN IN PARIS

Given my limitations, there's no way for me to have a reliably held belief about whether the probability of rain is 0.64475. But I can suspend on whether rain is 0.64475 likely for the reason that I can't have a reliably held belief. So the framework says that:

- It's *ex ante* rational for me to suspend judgement on whether rain is 0.67745 likely and *ex ante* irrational for me to believe it (by $\mathrm{RS}^{\mathrm{EA}}$ and $\mathrm{RB}^{\mathrm{EA}}$).

- If I suspend on whether rain is 0.67745 likely for the reason that I can't have a reliably held belief, then I suspend rationally (by $\mathrm{RS}^{\mathrm{EP}}$).

However, it is within my abilities to have a reliably held belief about whether rain is between 0.5-0.8 likely. So the framework says that:

- It's *ex ante* rational for me to believe that rain is between 0.5-0.8 likely, and *ex ante* irrational for me to suspend judgement on whether it's between 0.5-0.8 likely (by $\mathrm{RS}^{\mathrm{EA}}$ and $\mathrm{RB}^{\mathrm{EA}}$).

- If I come to believe that rain is between 0.5-0.8 likely in a reliable way, then my belief is rational (by $\mathrm{RB}^{\mathrm{EP}}$).

SuperMind, on the other hand, can have a reliably held belief about whether rain is exactly 0.67745. So, the framework says that:

- It's *ex ante* rational for her to believe that rain is exactly 0.67745 likely, and *ex ante* irrational for her to suspend judgement on whether it's exactly 0.67745 likely (by RS^{EA} and RB^{EA}).

- If she comes to believe that rain is exactly 0.67745 likely in a reliable way, then her belief is rational (by RB^{EP}).

Again, I think these are the right predictions.

§6. LIBERALISM AND ILLIBERALISM

The framework also makes some other interesting predictions.

§6.1. LIBERALISM

In one respect, the framework is very *liberal*. Suppose that my friend Morris is very cognitively limited. He rationally believes that p & q, but it is simply beyond his cognitive abilities to reliably infer that p by Conjunction Elimination, in the same way that it's beyond my cognitive abilities to have a reliably held belief about whether Theorem X is true, and so he suspends judgement on p. In that case, the framework says that Morris is rational in suspending judgement on p, even whilst believing p & q.

One might be unhappy with this result. *Intuitively*, one might say, he's irrational in suspending on p. I feel the pull of the intuition, but nevertheless I think it is the right result. If Morris suspends on p, it seems to me that he is thinking well *given his limitations*, just as I am when I suspend on Theorem X. He's like a person with very bad eyesight who suspends judgement on how many fingers the optician is holding up when all he sees is a blur.

The intuition that Morris is irrational in suspending on p is, I suspect, the product of projecting one's own standards and abilities on to him. *We* are able to reliably infer p from p & q, and hence rationally required to do so, and so we are tempted to think that *he* should be rationally required to do so as well. Tellingly, there is no such temptation in the Theorem X case. You're not able to reliably infer Theorem X from the axioms any more than I am, and so projecting your own standards and abilities does not yield the judgement that I'm irrational for not believing it.

But this kind of projection is illegitimate for two reasons. Firstly, it leads to an objectionable form of chauvinism. One of its consequences is that a community of mathematical geniuses, for whom inferring Theorem X from the axioms is child's play, would be correct in saying that *I* am irrational in suspending judgement on the Theorem, because *they* are rationally required to believe it. But that is surely wrong.[33] Secondly, it is obviously wrongheaded when applied to the moral domain. Just because the world's strongest man could lift the 400-kilo boulder under which my friend is currently pinned, and so would be morally obligated to rescue her, that does not mean that *I* am, because *I* cannot lift the boulder. It is hard to see why epistemology should be any different to morality in this respect.

On a different note, philosophers who write about non-ideal rationality sometimes say that cognitively non-ideal agents are not "ideally rational" – i.e., they are less rational than cognitively ideal agents (e.g., Richter (1990), Christensen (2007), Smithies (2015, 2023), Staffel (2019), Williamson (fc2)). I think that is a mistake. Morris isn't less *rational* than you and I when he suspends judgement on p, and you and I are no less rational than SuperMind when we suspend judgement on Theorem X, any more than the person with bad eyesight who suspends judgement on how many fingers the optician is holding up is less rational than people with good eyesight.

§6.2. ILLIBERALISM

Whilst the framework is very liberal in one respect, in another it is *illiberal:* it says that beliefs formed in unreliable ways are irrational even if you can't help it.

[33] I will return to this point in §7.3.

For example, suppose that you just cannot help but engage in egregious confirmation bias when it comes to certain topics. So, when you're thinking about these topics, you invariably give more weight to evidence that fits your prior beliefs than you do to evidence that goes against them. Suppose that this is just hardwired into you at a neurological level, so there's nothing you can do to stop it.

Nevertheless, on the plausible assumption that this is an unreliable way of forming beliefs,[34] the framework says that you're irrational when you come to believe things in this way. Again, I think this is the right result. You don't get to become rational just because it's not your fault that you're irrational.

So I think the framework makes some more good predictions here.

§6.3. OUGHT-IMPLIES-CAN

It is natural to think that a theory of non-ideal epistemic rationality must obey the maxim that "ought" implies "can". In other words, that you're not required to do something if you can't do it. How does this liberalism/illiberalism structure fit with that idea?

The first thing to say is that there are many possible ought-implies-can principles, corresponding to the many different interpretations of "ought" and "can". A theory of non-ideal epistemic rationality must surely obey some of them. But it is quite implausible that it must obey all of them. To take an obvious example, consider the "ought" of epistemic requirement, and an interpretation of "can" according to which one can φ only if φ-ing is compossible with one doing what one wants to do. As we've already seen, if "ought" implied "can" in this sense, then it would be fine for you to not believe that your friend is guilty of murder, despite overwhelming evidence, because you don't want to believe that he's guilty. But it isn't.

[34] Kelly (2008) and Dorst (2023) argue that some degree of confirmation bias can be epistemically rational. But they don't think that it is always epistemically rational. So imagine that you go way too far with it.

In virtue of its liberalism/illiberalism structure, my framework accepts this principle:

- If there are no nearby worlds where you have a reliably held belief about whether p, then you're not rationally required to believe that p.

But rejects this principle:

- If there are no nearby worlds where you don't believe that p, then you're not rationally required to not believe that p.

For the reasons given in §6.1 and §6.2, I think this is a plausible position to take on ought-implies-can.[35]

§7. AGAINST EPISTEMIC CONVENTIONALISM

§7.1. EPISTEMIC CONVENTIONALISM

So, that's my reliabilist theory of non-ideal rationality. Obviously, it's only a rough sketch – there are many details that need to be filled in.

But it's sufficient for my purposes here. Because I will now argue that even with this rough sketch, we can see that the theory provides us with a better take on non-ideal rationality than a recently popular theory, which I will call "conventionalism".

Notice that on my view what it's rational for cognitively non-ideal agents to believe depends entirely on their *individual* cognitive abilities and limitations. I'm not

[35] This take on ought-implies-can leaves it open that there might be epistemic dilemmas – situations where it's impossible for you to have a rational attitude towards p – as there might be circumstances where a person cannot have a reliably held belief about p but also cannot suspend on p for the reason that they can't have a reliably held belief. Elsewhere (Hughes 2019a, 2021, 2022, 2023, fc1, fc2; see also Hughes 2017, 2019b) I have argued that we should accept the possibility of epistemic dilemmas.

rationally required to believe have a belief about p because there are no nearby worlds where *I* have a reliably held belief. You are rationally required to believe that p because there are nearby worlds where *you* have a reliably held belief. Let's call this view "individualism":

> **INDIVIDUALISM:** What it's rational for non-ideal agents to believe depends entirely on their individual cognitive abilities and limitations.

Whether an individual has a reliably held belief about whether p in nearby worlds does not depend in any epistemologically interesting way on social conventions. So, individualism is a form of what I'll call "nonconventionalism":

> **NONCONVENTIONALISM:** What it's rational for non-ideal agents to believe is not determined by social conventions.

According to conventionalists, nonconventionalism is wrong.[36] The conventionalist argues that it isn't *individual* abilities and limitations that determine the standards of non-ideal rationality. Rather, they endorse:

> **CONVENTIONALISM:** What it's rational for non-ideal agents to believe is determined by social conventions.

How do social conventions determine non-ideal rationality? Dogramaci (2015) argues that within our communities we have conventions whereby we tacitly endorse certain sets of truth-conducive belief-forming rules and reject others, and call beliefs formed by the endorsed rules "rational" and beliefs formed by the rejected rules "irrational". It is these conventions, he argues, that fix what it is rational and irrational for cognitively non-ideal agents to believe. On this view we should reject individualism in favour of "communism":

[36] Who is a conventionalist? Dogramaci (2015) comes closest, but he doesn't discuss all the arguments for conventionalism I'll look at. Lasonen-Aarnio (ms) discusses some of the data I'll take to motivate conventionalism in a way that seems sympathetic to it, though she doesn't explicitly endorse the view. Carr (2022) discusses the view at length, but only argues for a conditional: *if* her description of the data is correct, *then* conventionalism is the best explanation of it. So, there may not be any one person who accepts everything I'll say about the view. Nevertheless, I expect many epistemologists will find it appealing.

> **COMMUNISM:** What it's rational for non-ideal agents to believe depends on the standards endorsed by the community.

On the communist view, it isn't random which belief-forming rules a community endorses and which it rejects (Dogramaci 2015, Carr 2022). Within a community we need to be able to trust one another's testimony, in order to pool and share information. However, we are only able to trust a person's testimony when they form their beliefs using the same belief-forming rules as us. As a result, the community must coordinate on a shared set of accepted belief-forming rules, otherwise, it will not be able to pool and share information. Various pressures, such as the need for most of the members of the community to be able to follow the rules, and perhaps even evolutionary pressures, will result in the set of endorsed rules being a rather small proper subset of all the truth-conducive belief-forming rules.[37]

But even though it's not random which rules we endorse and which we reject, going with our particular set of rules is merely conventional in the sense that (1.) We could have endorsed a different set of truth-conducive rules. (2.) If we had endorsed a different set of rules, then those would have been the rational ones to use. (3.) The set of rules we've settled on is in some sense no better than any other set of truth-conducive rules. (4.) It is to some extent arbitrary that we've settled on this set of rules.

The communist story isn't mandatory for the conventionalist. One might offer some other story about how social conventions fix non-ideal rationality. But it is a natural complement to conventionalism, and it will play a role in the arguments for conventionalism I will now discuss.

Should we be conventionalists? I don't think so. I will look at three arguments for conventionalism in the literature. In the course of the discussion, we will get a better sense of the conventionalist view. In each case, I'll present the argument and

[37] Relatedly, Hannon and Woodard (fc) argue that communism best explains why epistemic norms have normative force even in cases where in our practical interest to violate them.

then argue that my nonconventionalist view explains the data better than conventionalism.

§7.2. THE ARGUMENT FROM ALETHIC TIES

I'll call the first argument, which is put forward by Dogramaci (2015) and sympathetically discussed by Carr (2022), the "argument from alethic ties".

It starts with the observation that there are many equally truth-conducive belief-forming rules ("truth-conducive" in the sense that when you employ the rule you won't be taken from a truth to a falsehood). For instance, consider the following "easy" rules and "hard" rules:

EASY RULES

Infer Q from P and P → Q **(Modus Ponens)**
Infer P from P & Q **(Conjunction Elimination)**
Infer P or Q from P **(Disjunction Introduction)**

HARD RULES

Infer Fermat's Last Theorem straight from the axioms **(Fermat's Rule)**
Infer Lagrange's Four-Square Theorem straight from the axioms **(Lagrange's Rule)**
Infer Peirce's Law from no premises whatsoever **(Peirce's Rule)**

Whilst we can rationally use the easy rules to form new beliefs, intuitively, it's irrational for us to use (or rather, try to use) the hard rules: if you look at the axioms of arithmetic and immediately conclude from them that Fermat's Last Theorem ("FLT") must be correct, with no intermediary inferential steps, then intuitively your belief isn't rational.

On the face of it, conventionalists say, this is puzzling. Given that all the rules are equally truth-conducive, why should it only be rational to use the easy rules and not the hard ones too? The conventionalist argues that the best explanation is that as a community we've settled on tacitly endorsing the easy rules and tacitly banning the hard ones as acceptable belief-forming rules, in order to facilitate the pooling and sharing of information. Hence, the conventionalist takes the contrast between the two sets of rules to speak in favour of conventionalism.

§7.3. A NONCONVENTIONALIST EXPLANATION OF THE DATA

I'm not persuaded by this argument.

Take FERMAT'S RULE. Conventionalists are surely correct that it's not rational for the average person to look at the axioms of arithmetic and immediately conclude from them that FLT must be correct, with no intermediary inferential steps. But my framework can easily explain why this is without appealing to conventionalist ideas. Because whilst FERMAT'S RULE is logically valid, it doesn't follow that anyone who forms a belief in accordance with it has a reliably formed belief. Whether the belief is reliably formed depends on the *way* that it was formed. And the only way for the average person to move straight from the axioms to FLT with no intermediary steps is by guessing. But guessing is a paradigmatically unreliable way of forming beliefs.[38]

So, when it comes to the average person, my nonconventionalist framework accounts for the data just as well as conventionalism.

[38] When I've given talks on this, some audience members have questioned how a belief formed in accordance with FERMAT'S RULE could be unreliably formed, given that the rule is valid, and hence perfectly reliable. So let me say a bit more. My answer is that we need to distinguish between the reliability of the rule *itself* and the reliability of the *way* the person who forms a belief in accordance with it forms the belief. One can act in accordance with a rule without acting in a way that reliably gets you the goods that the rule reliably gets you. Consider the rule 'You should: if there are ticket inspectors on the train, have a valid ticket' and the good of not getting fined. This rule is perfectly reliable in the sense that whenever you act in accordance with it you won't get fined. But now imagine that you don't buy a ticket, but, luckily for you, there are no ticket inspectors on your train. You act in accordance with a perfectly reliable rule, but clearly the way that you go about things does not reliably result in you not getting fined. Similarly, if you leap straight from the axioms to FLT, with no intermediary steps, then even though you form your belief in accordance with a perfectly reliable rule (FERMAT'S RULE), the way that you form your belief does not reliably result in you having true beliefs or knowledge because, given your cognitive limitations, this cannot be anything other than a guess, and in many other cases this way of forming beliefs would result in you having false beliefs. (I'm assuming here that ways of forming beliefs should be individuated in a somewhat coarse-grained way, so that the way you form your belief is best described as 'guessing'. That seems very plausible to me).

But what happens when we look beyond the average person? Here, I think the data favours my nonconventionalist approach over conventionalism.

With that in mind, I'd like to consider four more scenarios. The first involves people with very low cognitive acuity living in a community of people with average cognitive acuity. The second involves people with average cognitive acuity living in a community of people with very low cognitive acuity. The third involves people with average cognitive acuity living in a community of people with very high cognitive acuity. The fourth involves people with very high cognitive acuity living in a community of people with average cognitive acuity. In each scenario, the protagonist forms a doxastic attitude that is at odds with the epistemic practices of the community. According to conventionalism, the protagonist's attitude should be irrational. According to my framework, it is rational. I will suggest that, in each scenario, my framework delivers the more plausible verdict.

§7.3.1. Low Acuity, Average Acuity Community

When it comes to people with very low cognitive acuity living in an average acuity community, we're talking about people like Morris.

As we've already seen, according to the reliabilist framework, if Morris is incapable of reliably inferring p from p & q, then it's rational for him to suspend judgement on p. This is a result of the framework's liberalism (§6.1). As I've already said, I think that's the right result. He's thinking well *given his cognitive limitations*. He's like the person with bad eyesight who suspends judgement on how many fingers the optician is holding up. But proponents of conventionalism will have to maintain that Morris is irrational if he suspends judgement on p, because reasoning by Conjunction Elimination has the stamp of approval in the average community. So, conventionalism makes the wrong prediction here, whereas my nonconventionalist framework makes the right prediction.

§7.3.2. Average Acuity, Low Acuity Community

Now change the case. Instead of imagining Morris in our community, imagine that you grew up in a community of Morrises. You believe that p, having reliably inferred it from p & q. But to the people around you this looks like some kind of magic. For them, inferring p from p & q is as difficult as inferring FLT straight from the axioms is for us. They think you're irrational. Are they right? Surely not. But according to conventionalism, they should be, because Conjunction Elimination is not an endorse rule of belief-formation in the community. By contrast, my framework predicts that they are wrong. So, conventionalism makes the wrong prediction here, whereas my nonconventionalist framework makes the right one.

§7.3.3. Average Acuity, Exceptional Acuity Community

Next, consider a person with average cognitive acuity living in a community of people with very high cognitive acuity.

So, imagine that you grew up in a community of mathematical geniuses; the kind of people who can see that FLT follows from the axioms as easily as you can see that p follows from p & q. Unfortunately, none of their brilliance has rubbed off on you.

Now, suppose that they ask you: is FLT a theorem? You suspend judgement, because you have no idea – figuring it out is far beyond your abilities.[39] Are you irrational? Maybe all the people around you think you are. "Can't you just *see* whether FLT is a theorem?" they ask you. You reply: "No, I can't see that. And given that I can't, it's rational for me to suspend judgement. If I were to form a belief about it, I'd just be guessing".

As I said in §6.1, that seems to me to be a perfectly reasonable reply. But proponents of conventionalism are going to have to say that it's wrong. They'll have to say that you're irrational, because FERMAT'S RULE has the stamp of approval in the

[39] Assume that you haven't heard of Andrew Wiles's proof of FLT.

community. So, again, conventionalism makes the wrong prediction here whereas my nonconventionalist framework makes the right one.

§7.3.4. Exceptional Acuity, Average Acuity Community

Finally, let's consider people with exceptional cognitive acuity living in a community of people with average cognitive acuity.

According to my framework, if a person in an average community can reliably infer FLT straight from the axioms, then it's rational for them to do so. But according to conventionalism it's irrational, because FERMAT'S RULE is banned in the average community.

Who's right? One conventionalist, Dogramaci (2015), appeals to BonJour's (1980) clairvoyance cases to argue that it's irrational. Like BonJour, Dogramaci maintains that Norman the clairvoyant's belief that the President is in NYC isn't rational even though it is reliably formed. If so, he argues, neither is the belief of the person with exceptional cognitive acuity who reliably infers FLT straight from the axioms.

There is not enough space to get into the complexities of the Norman case here, but let me briefly describe the weaknesses of this argument.

The argument depends on three premises: (1.) The person with exceptional cognitive acuity who infers FLT straight from the axioms must be relevantly like Norman. (2.) Norman's belief is reliably formed. (3.) Norman's belief is irrational. All three premises are highly questionable.

Starting with the second premise: there are reasons to think that Norman's belief isn't reliably formed. It may be that, on the best way of developing reliabilism, Norman's belief isn't reliably formed because the *global* cognitive dispositions he manifests are not conducive to getting it right; the way that BonJour describes Norman suggests that he would believe that the President is in NYC even if the *local* mechanism that produces the belief did not yield a high ratio of true to false beliefs. If that's right, then we can simply stipulate that the global cognitive dispositions that the person who infers FLT straight from the axioms manifests are conducive to getting it right, and so that they would not have believed FLT if the local mechanism that produces the belief had been unreliable. If so, premise (1) of Dogramaci's argument is false: the person is not relevantly like Norman.

On the third premise: there are also reasons to doubt that Norman's belief is irrational. Firstly, if we deny that Norman's belief is rational, then it will be very difficult to maintain that infants and non-human animals are capable of having rational beliefs, as their situations – they have reliable faculties but no evidence for or against the claim that those faculties are reliable – appear to be precisely analogous to Norman's (Kornblith 2012). That is a very unwelcome result. Secondly, the Norman case is an early instance of the now-familiar phenomenon of apparent defeat by higher-order evidence. But one of the lessons of the recent literature on higher-order evidence is that *every* theory of rationality will face higher-order evidence worries, where one's belief satisfies the theory's conditions on rationality, but one lacks evidence that it satisfies them, or even has evidence indicating that it doesn't satisfy them (Lasonen-Aarnio 2014). So unless we want to give up on the theory of rationality entirely, we should resist the idea that the people in these sorts of these cases are irrational. But if Norman's belief is rational after all, then even if premise (1) is true – that is, even if the person who infers FLT straight from the axioms is relevantly like Norman – Dogramaci's argument fails to show that they are irrational.

I won't take a stance here on which of these responses to Dogramaci's argument is ultimately best. But between them, they strongly suggest that the argument is unsound.

Furthermore, moving away from clairvoyance cases, I think that another analogy with vision favours my view over conventionalism here. To see this, imagine that you have extraordinarily good eyesight – you can reliably identify small objects at very large distances. Far better than any normal person. Out on a walk one day you clearly see a rabbit in the far distance (say, two miles away).[40] Are you irrational for believing that there's a rabbit over there? Surely not; you can see it.

[40] Eagles can identify rabbits from two miles away (Grambo 1999).

And it is still rational for you to believe it *even if* there is widespread skepticism in your community about your ability. But by parity of reasoning, it seems that conventionalists must say that you are irrational. Just as you can see things that others in your community can't, so too can the person who can reliably infer FLT straight from the axioms. So if it's irrational for that person to believe FLT, then it should also be irrational for you to believe that there's a rabbit over there. But that is surely incorrect.

Summing up on the argument from alethic ties, I think that my nonconventionalist reliabilist framework does just as good a job at explaining the data that the conventionalist appeals to here and does a better job of explaining other data in the vicinity.

§7.4. THE ARGUMENT FROM ARBITRARY BAR-LOWERING

The second argument for conventionalism is discussed by Carr (2022). I'll call it the "argument from arbitrary bar-lowering". It goes like this. Consider the following two lists of cognitive limitations:

LIST A

- Limited computational power
- Limited storage space
- Limited processing speeds
- Integration of different cognitive systems
- Imperfect information retention

LIST B

- Implicit bias
- Delusional reasoning
- Unreliable heuristics (e.g., Representativeness)
- Misinterpreting statistical phenomena as having causal explanations
- An inflated sense of one's own driving ability

- Over-optimism and over-pessimism
- Overestimating the moral superiority of one's own side in a disagreement

Carr observes that non-ideal epistemologists usually think there is a difference between these two lists. In particular, they usually think that the limitations on List A "lower the bar" for non-ideal rationality, whereas the limitations on List B don't. That is: doxastic attitudes that result from the influence of the limitations on List A can be rational, but those that result from the influence of the limitations on List B can't be rational.[41,42]

For example, I suspend judgement on whether Theorem X is true. And I do so because I don't have enough computational power to figure it out. Plausibly, my suspending is rational. This supports the idea that limitations on computational power belong on List A – limitations that lower the bar for rationality.

By contrast, suppose that you believe that p because you engage in egregious confirmation bias: you want to believe that p, so you give far more weight to evidence confirming it than you do to evidence disconfirming it. In this case, the influence of your cognitive limitation doesn't make your belief rational – it's irrational, given the way it was formed. This supports the idea that confirmation bias belongs on List B – limitations that don't lower the bar for rationality.

So, that's the idea. Now, Carr maintains that there's no *principled* reason why the limitations on List A should lower the bar, but not the limitations on List B. Rather, the best explanation of the difference, she argues, is that as a community we have adopted a social convention whereby we accept the influence of the limitations on List A, but prohibit the influence of the limitations on List B. In connection with this she writes: "Given that we have a large array of cognitive imperfections and can't continually mitigate all of them, it makes sense to implement some conventions determining which to try to mitigate and which to accept as our fate. We then treat the former as constituting irrationality and the latter as lowering the

[41] More precisely: the limitations on List A shrink the modal base used to determine the truth-conditions of statements of the form "S's doxastic attitude towards p is rational", but the limitations on List B don't.

[42] Carr doesn't name names. She only attributes the idea that only the items on List A lower the bar to non-ideal epistemologists generically. But two recent examples are Thorstad (fc) and Lasonen-Aarnio (ms).

bar for rationality" (Carr 2022: 1151). Hence, the difference between Lists A and B is, she argues, evidence that conventionalism is true in non-ideal epistemology.

§7.5. A NONCONVENTIONALIST REPLY

I'm not persuaded by this argument either. It seems to me that those who see a difference between List A and List B have the data wrong. I don't think it's true that the limitations on List A straightforwardly lower the bar whereas the limitations on List B don't. Rather, if my judgements about cases are on the right track, then how cognitive limitations make a difference to the rational standing of your beliefs depends on the way they operate in belief-formation.

In particular, if you suspend judgement on p because it's beyond your cognitive abilities to have a reliably held belief about it, then, as I've said, it seems to me that you're rational (consider Morris suspending on p, or me suspending on Theorem X, for example). Non-ideal rationality is liberal. But if, owing to your cognitive limitations, you come to believe that p in an unreliable way, then your belief is irrational (consider the person who cannot help but engage in confirmation bias, for example). Non-ideal rationality is also illiberal.

Crucially, this distinction doesn't track the taxonomy on offer with List A and List B. To see this, take one of the items on List A: limited computational power. The suggestion is that it lowers the bar for non-ideal rationality. So, attitudes formed as a result of limited computational power are rational. But suppose that your limited computational power causes you to form a belief that p by guessing. In that case, your belief is irrational, even though the limitation is on List A: limitations that allegedly lower the bar for rationality.

Now take one of the items on List B: confirmation bias. The suggestion is that it doesn't lower the bar for rationality. So, attitudes formed as a result of the bias are irrational. But suppose that you suspend judgement on whether p because you know that if you were to have an opinion about it, that opinion would be unreliably formed because you'd engage in confirmation bias in arriving at it. In that case, surely you're rational in suspending on p, even though the limitation is on List B – limitations that allegedly don't lower the bar for rationality.

According to the argument from arbitrary bar-lowering, there's no principled reason why the limitations on List A should lower the bar but not the limitations on List B, and the best explanation of the difference is that what it's rational for non-ideal agents to believe is determined by social convention. But if what I've just said is right, then proponents of the argument are wrong about the data that needs to be explained in the first place, because it's simply not true that the limitations on List A lower the bar whilst those on List B don't. And so there's nothing here that we need conventionalism to explain.

Why, then, does it initially seem plausible that the limitations on List A lower the bar but not the limitations on List B? The explanation might be that when we think about the limitations on List A we tend imagine cases where they cause us to suspend judgement, whereas when we think about the limitations on List B we tend to imagine cases where they cause us to form beliefs. Beliefs caused by the limitations on List B are indeed usually irrational, and suspending judgement owing to the limitations on List A is indeed usually rational. But once we appreciate the fact that the influence of limitations on both lists can lead to both belief *and* suspension, the illusion that there is a neat taxonomy here disappears.

§7.6. THE ARGUMENT FROM IRRELEVANT OF INDIVIDUAL VARIATIONS IN ABILITY

The third argument for conventionalism is also found in Carr (2022). I'll call it the "argument from the irrelevance of individual variations in ability".

The argument starts with the observation that the standards of non-ideal rationality are not straightforwardly sensitive to individual variations in ability. In particular, some beliefs, like those of somatoparaphrenics and conspiracy theorists are irrational even though the people who hold them can't do any better.

The conventionalist argues that this would be surprising if the standards of non-ideal rationality aren't fixed by coordination-promoting social conventions. After all, other normative domains that (arguably) aren't conventional, like morality, are straightforwardly sensitive to individual variations in ability – you're not obligated to give $100 billion to charity, but Jeff Bezos is. Why should epistemology be any different?

By contrast, the argument goes, insensitivity to individual variations in ability exactly what we would expect if the standards are fixed by coordination-promoting social conventions, because "Coordination in epistemic standards would be near impossible to achieve if the applicability of epistemic evaluations depended on knowledge of idiosyncratic psychological features of individual believers" (Carr 2022: 1154) So, the conventionalist argues, the irrelevance of individual variations in ability is evidence for conventionalism.

§7.7. A NONCONVENTIONALIST REPLY

Again, I'm unpersuaded. In this case, I agree with the characterisation of the data. The beliefs of somatoparaphrenics and conspiracy theorists are irrational even though they can't do any better. But I don't think this is evidence for conventionalism over nonconventionalism, because my nonconventionalist framework also predicts the data.

This is where the illiberalism of the framework comes into play again. As I pointed out, the framework is illiberal in the sense that it says that unreliably formed beliefs are irrational even if you can't help it. The example I used was confirmation bias. Even if the bias is hardwired into you at a neurological level, and so there's nothing you can do about it, the framework says that your resulting beliefs are irrational because they are unreliably formed. But exactly the same goes for somatoparaphrenic and conspiratorial reasoning. These are paradigmatic cases of unreliable belief formation, so the framework says that they're irrational, even if the people engaging in them can't help it. So, nonconventionalism predicts the data just as well as conventionalism here.

We've looked at three arguments for conventionalism. In each case, my nonconventionalist framework handles the data at least as well, or better, than conventionalism.

§8. CONCLUSION

Summing up, although I have only sketched the outlines of a reliabilist theory of non-ideal rationality here, it seems to me that the theory is a promising one and that if the theory is on the right track, then we should reject conventionalism in favour of nonconventionalism.[43]

§9. REFERENCES

Ballantyne, N. (2019). *Knowing Our Limits*. Oxford University Press

Beddor, B. & Pavese, C. (2020). "Modal Virtue Epistemology" *Philosophical and Phenomenological Research* 101 (1): 61-79

Begby, E. (2021). *Prejudice: A Study in Non-Ideal Epistemology.* Oxford University Press

BonJour, L. (1980). "Externalist Theories of Empirical Knowledge" *Midwest Studies in Philosophy* 5 (1): 53-73

Cherniak, C. (1990). *Minimal Rationality*. MIT Press.

Christensen, D. (2007). "Does Murphy's Law Apply in Epistemology? Self-Doubt and Rational Ideals" Gendler, T. & Hawthorne, J. (eds.) *Oxford Studies in Epistemology*. Oxford University Press.

Carr, J. (2022). "Why Ideal Epistemology?" *Mind* 131 (524): 1131-1162

Cohen, S. (1984). "Justification and Truth" *Philosophical Studies* 46 (3): 279-295

[43] For helpful discussion, thanks to Jennifer Carr, Daniel Kodsi, Maria Lasonen-Aarnio, Tim Williamson, and audiences at the University of Gothenburg, the University of Helsinki, the University of Leeds, the University of Oslo, Oxford University, Umea University, Freie University Berlin, The Inland Norway University of Applied Sciences, and the Newington Green Seminar.

Conee, R. & Feldman, R. (1998). “The Generality Problem for Reliabilism” *Philosophical Studies* 89 (1): 1-29

DiPaolo, J. (2019). “Second Best Epistemology: Fallibility and Normativity” *Philosophical Studies* 176: 2043-2066

Dogramaci, S. (2015). “Communist Conventions for Deductive Reasoning” *Nous* 49 (4): 776-799

Dorst, K. (2020). “Evidence: A Guide for the Uncertain” *Philosophy and Phenomenological Research* 100 (3): 586-632

Dorst, K. (2023). “Rational Polarization” *Philosophical Review* 132 (3): 355-458

Dretske, F. (1981). *Knowledge And The Flow of Information*. MIT Press.

Dusi, G. (Forthcoming). “Reliabilist Epistemology Meets Bounded Rationality”. *Synthese*

Gao, J. (2019). “Credal Pragmatism” *Philosophical Studies* 176: 1595-1617

Gao, J. (2023). “Should credence be sensitive to practical factors? A cost-benefit analysis” *Mind & Language* 38 (5): 1238-1257

Garber, D. (1983). “Old Evidence and Logical Omniscience in Bayesian Confirmation Theory” Earman, J. (ed.) *Testing Scientific Theories*. University of Minnesota Press

Gigerenzer, G, *et al*. (1999). *Simple Heuristics That Make Us Smart*. Oxford University Press

Gigerenzer, G & Selten, R. (2002). *Bounded Rationality: The Adaptive Toolbox*. MIT Press.

Gigerenzer, G. (2021). “Axiomatic Rationality and Ecological Rationality” *Synthese* 198: 3547-3564

Goldman, A. (1979). “What Is Justified Belief?” Pappas, G. (ed.) *Justification and Knowledge*. Dordrecht

Goldman, A. (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard University Press

Goldman, A. (2021). “A Different Solution to the Generality Problem for Process Reliabilism” *Philosophical Topics* 49 (2): 105-111

Graham, P. (2017). “Normal Circumstances Reliabilism: Goldman on Reliability and Justified Belief” *Philosophical Topics* 45 (1): 33-61

Grambo, R. (1999). *Eagles*. Voyageur Press.

Hacking, I. (1967). “Slighty More Realistic Personal Probability” *Philosophy of Science* 34 (4): 311-325

Hannon, M. & Woodard, E. (Forthcoming). “The Construction of Epistemic Normativity” *Philosophical Issues*

Hirvela. J. (Manuscript). “A Modal Theory of Justification”

Hughes, N. (2017). “No Excuses: Against the Knowledge Norm of Belief” *Thought: A Journal of Philosophy* 6 (3): 157-166

Hughes, N. (2019a). “Dilemmic Epistemology” *Synthese* 196: 4059-4090

Hughes, N. (2019b). “Uniqueness, Rationality, and the Norm of Belief” *Erkenntnis* 84: 57-75

Hughes, N. (2021). “Who’s Afraid of Epistemic Dilemmas?” McCain, K., Stapleford, S., & Steup, M. (eds) *Epistemic Dilemmas: New Arguments, New Angles*. Routledge

Hughes, N. (2022). “Epistemology Without Guidance” *Philosophical Studies* 179: 163-196

Hughes, N. (2023). “Epistemic Feedback Loops (Or: How Not to Get Evidence)” *Philosophy and Phenomenological Research* 106 (2): 368-393

Hughes, N. (2024). “Evidence and Bias” Lasonen-Aarnio, M. & Littlejohn, C. (eds.) *The Routledge Handbook of the Philosophy of Evidence*. Routledge.

Hughes, N. (Forthcoming-1). “Epistemic Dilemmas Defended” Hughes, N. (ed.) *Essays on Epistemic Dilemmas*. Oxford University Press.

Hughes, N. (Forthcoming-2). “Epistemic Dilemmas: A Guide” Hughes, N. (ed.) *Essays on Epistemic Dilemmas*. Oxford University Press.

Icard. T. (Manuscript). *Resource Rationality*.

Ichikawa, J. (2022). “You Ought to Have Known: Positive Epistemic Nomrs in a Knowledge-First Framework” *Synthese* 200 (5): 1-23

Kearl, T. & Willard-Kyle, C. (Forthcoming). “Epistemic Cans” *Philosophy and Phenomenological Research*.

Kelly, T. (2008). “Disagreement, Dogmatism, and Belief Polarization” *Journal of Philosophy*. 105 (10): 611-633

Kornblith, H. (2012) *On Reflection*. Oxford University Press.

Kratzer, A. (1977). “What ‘Must’ and ‘Can’ Must and Can Mean” *Linguistics and Philosophy* 1 (3): 337-355

Lasonen-Aarnio, M. (2010). “Unreasonable Knowledge” *Philosophical Perspectives* 24 (1): 1-21

Lasonen-Aarnio, M. (2014). “Higher-Order Evidence and the Limits of Defeat” *Philosophy and Phenomenological Research* 88 (2): 213-345

Lasonen-Aarnio, M. (2020). “Enkrasia or Evidentialism? Learning to Love Mismatch” *Philosophical Studies* 177 (3): 597-632

Lasonen-Aarnio, M. (2021). “Dispositional Evaluations and Defeat” Brown, J. & Simion, M. (eds.) *Reasons, Justification, and Defeat*. Oxford: Oxford University Press

Lasonen-Aarnio, M. (Forthcoming-1). “Virtuous Failure and Victims of Deceit” Dutant, J. & Dorsch, F. (eds.) *The New Evil Demon Problem*. Oxford University Press

Lasonen-Aarnio, M. (Forthcoming-2). “Perspectives and Good Dispositions” *Philosophy and Phenomenological Research*

Lasonen-Aarnio, M. (Manuscript). *Feasible Dispositions: A Normative Framework*

Lewis, D. (1976). “The Paradoxes of Time Travel” *American Philosophical Quarterly* 13 (2): 145-152

Littlejohn, C. (2013), "The Russellian Retreat" *Proceedings of the Aristotelian Society* 113 (3): 293–320

Littlejohn, C. (Forthcoming). “A Plea for Epistemic Excuses” Dutant, J. & Dorsch, F. (eds.) *The New Evil Demon Problem*. Oxford University Press

Lyons, J. (2019). “Algorithm and Parameters: Solving the Generality Problem for Reliabilism” *Philosophical Review* 128 (4): 463-509

Morton, A. (2012). *Bounded Thinking: Intellectual Virtues for Limited Agents*. Oxford University Press

Moss, S. (2018). “Moral Encroachment” *Proceedings of the Aristotelian Society* 118 (2): 177-205

Nelson, M. (2010), “We Have No Positive Epistemic Duties” *Mind* 119 (473): 83–102

Pettigrew, R. (2021). “Logical Ignorance and Logical Learning” *Synthese* 198: 9991-10020

Richter, R. (1990). “Ideal Rationality and Hand Waving” *Australasian Journal of Philosophy* 68 (2): 147-156

Rysiew, P. (2008). “Rationality Disputes – Psychology and Epistemology” *Philosophy Compass* 3 (6): 1153-1174

Simion, M. (2024a). “Resistance to Evidence and the Duty to Believe” *Philosophy and Phenomenological Research* 108 (1): 203-216

Simion, M. (2024b). *Resistance to Evidence*. Cambridge University Press

Singer, D. (2023). *Right Belief and True Belief*. Oxford University Press.

Skipper, M. & Bjerring, J. (2020). ”Bayesianism for Non-Ideal Agents” *Synthese* 87: 93-115

Smithies, D. (2015). “Ideal Rationality and Logical Omniscience” *Synthese* 192 (9): 2769-2793

Smithies, D. (2022). “The Epistemic Function of Higher-Order Evidence” Silva, P. & Oliveira, L. (eds.) *Propositional and Doxastic Justification: New Perspectives*. Routledge

Staffel, J. (2019). *Unsettled Thoughts: A Theory of Degrees of Rationality*. Oxford University Press

Sutton, J. (2007). *Without Justification*. MIT Press

Thorstad, D. (Forthcoming). “Why bounded rationality (in epistemology)?” *Philosophy and Phenomenlogical Research*

Thorstad, D. (Manuscript). *Inquiry Under Bounds*

Williamson, T. (2000). *Knowledge and Its Limits*. Oxford University Press

Williamson, T. (Forthcoming-1). "Justifications, excuses, and sceptical scenarios" Dutant, J. & Dorsch, F. (eds.) *The New Evil Demon Problem*. Oxford University Press

Williamson, T. (Forthcoming-2). "Knowledge, Credence, and the Strength of Belief” Floweree, A. & Reed, B. (eds.) *Expansive Epistemology: Norms, Action, and the Social World*. Routledge