Robustness and the Distinctive Properties of Epistemic Ideals

Marc-Kevin Daoust, École de technologie supérieure
Forthcoming in *Australasian Journal of Philosophy*

Why care about ideal epistemic norms? Why not merely care about norms that agents like us can actually meet? In this paper, I make two claims. First, I argue that, if we want robust epistemic norms, we can't just do idealised epistemology. We need to confirm that the results and observations that are central in idealised epistemology also obtain in other contexts. Second, I argue that, if we really want to capture the fundamental nature of epistemic normativity, we need the sort of frameworks that come from non-ideal epistemology. I develop this argument in response to ideas put forward by Jennifer Carr (2022) surrounding the distinctive properties of epistemic ideals.

Keywords: ideal rationality, bounded rationality, second-best epistemology, norms of perfection, constrained optimisation

1 Introduction

In the past decades, many disciplines have had important internal debates about the place of normative ideals. For instance, in welfare economics, many theorists have argued that we should not always approximate ideals. This is the general lesson behind the Theory of the Second Best (Wiens 2020; Lipsey and Lancaster 1956). In political philosophy, the distinction between ideal and nonideal theory, and the relevance of political ideals, is an important concern among theorists (Valentini 2012). In ethics, some philosophers wonder whether it is wise to follow the guidance of ideal advisors (Smith 1994; Sampson 2022).

These debates now seem to be moving into epistemology. In the past couple of years, we see a surge of publications on this topic. Think, for instance, of DiPaolo's (2019) Second-Best Epistemology, Staffel's (2020) approximation model of the Bayesian ideals, McKenna's (2023) argument in favour an epistemology suited for the real world, and Carr's (2022) justification of epistemic ideals. And it's a good thing that we're finally having these discussions, since there are a lot of ideals in epistemology. Consider the Bayesian frameworks that ignore an agent's limited cognitive capacities. On these accounts, agents should be probabilistically coherent, update their

credences in accordance with Conditionalization, satisfy some inter-level coherence requirements like Immodesty or Reflection principles, and satisfy various others principles like the Principal Principle and the Principle of Indifference (Carr 2022, 2; Pettigrew 2016). Perfect epistemic cherubs can comply with these requirements, but imperfect and cognitively limited agents like us can't. So, why care about these ideals?

Bayesians have different options for justifying the ideals they describe. For instance, they could turn to models of ideal advisors (Smith 1994; Sampson 2022; Karlan 2021). They can argue that epistemic ideals matter because we are in a position to approximate them (Staffel 2020). They can say that the Bayesian ideals operate like the values of a function that we want to maximise, even if we can't fully maximise all the values of the function (Christensen, 2007: 24; 2010). And of course, there could be other justifications of the Bayesian ideals.

In this paper, I make two points. They are both related to the concept of robustness. First, I argue that, if we want robust epistemic norms, we can't just do ideal epistemology. The observations made in ideal epistemology rest on a number of simplifying, false assumptions. This can generate skepticism regarding these claims. In order to overcome this skepticism, we can do robustness analysis, and confirm that these observations can also be derived from more realistic models. However, this type of robustness analysis is not possible without models of bounded rationality.

Second, I argue that, if we want to capture the fundamental nature of epistemic normativity, we need the sort of models that come from non-ideal epistemology. In order to make this point, I first respond to a recent justification of epistemic ideals. According to Jennifer Carr (2022), only ideal epistemic norms are normatively robust, in the sense that they can capture the fundamental nature of epistemic normativity. Call this the Exclusivity Argument. The Argument rests on three putative distinctive properties of epistemic ideals. However, the distinctions between ideal and non-ideal epistemic norms underlying the Exclusivity Argument are implausible. Also, if we take her argument seriously, non-ideal epistemology might be a more suitable way to capture the fundamental nature of epistemic normativity. This is so, because certain brands of non-ideal epistemology can remain neutral on many contextual factors, such as the constraints that agents face, while their idealised counterparts do not remain neutral on these contextual factors. Ignoring constraints is a contextual assumption.

The paper goes as follows. I first offer a robustness argument in favour of non-ideal epistemology (§2). Then, I clarify the Exclusivity Argument, which relies on a different but related notion of robustness (§3). I present some initial concerns regarding the Exclusivity Argument (§§4-5). Put together, these worries lead me to my second, main argument. Models from non-ideal epistemology, such as constrained optimisation, are better suited to capture the fundamental nature of epistemic normativity (§6).

Before we start, a quick methodological remark: There are many ideal/non-ideal distinctions in philosophy.[1] Accordingly, when we talk about ideal and non-ideal epistemology, there is a risk of equivocation. Perhaps the distinction brings more trouble than clarity, and should be *abandoned* rather than conceptualised. However, since I wish to address the Exclusivity Argument, I will not question the distinction between ideal and non-ideal epistemology. I will also assume, like many authors, that the standard Bayesian norms (probabilism, Conditionalization, etc.) are requirements of ideal rationality.

2 Robustness and Ideal Epistemology

In this section, I argue that, if we want to identify robust results in epistemology, we need more than idealised theories and models. In other words, robustness gives us a reason think that, if we want to do good ideal epistemology, then we need its non-ideal counterpart.

A classic problem, in philosophy of science, concerns the degree of confidence we should have in the predictions and observations derived from highly simplified models. Many of the phenomena scientists are interested in are exceedingly complex. Think, for example, of biodiversity, economic growth or the impact of public policies. It's hard to elaborate complete, accurate and general models of these phenomena. The models available for understanding these phenomena are often simplified and based on false assumptions. For instance, some models for predicting the behavior of ocean waves assume that the ocean floor has infinite depth (Fuller and Schulz 2021: 6-7). Of course, this assumption is false. The seafloor has a finite, irregular depth.

Suppose a model generates a conclusion C. We know that models rely on false assumptions. Accordingly, we can be skeptical of C. Is C an artifact of the model's inaccurate assumptions? Can we trust C in real-life contexts, that is, in contexts where some of the model's assumptions do not obtain? One method we can use to overcome this skepticism is robustness analysis. The method consists

1 See Valentini (2012) and Daoust (2021a; forthcoming).

of analysing whether conclusion C can be derived from other, independent models. Again, consider the ocean waves problem. We could check whether C is also a consequence of models that make distinct assumptions about the depth of the ocean floor. If we can derive C from models that assume that ocean floors are shallow, deep and infinite, then it seems that C is not an artifact of the false assumption that the ocean floor has infinite depth.

There is a detailed literature on what we learn, exactly, from robustness analysis. Some merely say that this type of analysis allows us to rule out the possibility that a conclusion depends on the model's artificial assumptions.[2] Others think that robustness analysis has other benefits regarding, among other things, the confirmation of theorems,[3] the confirmation of scientific theories,[4] or the identification of causal relations.[5] Here I merely make the weak assumption that robustness analysis can alleviate some types of skepticism.

I contend that this argument extends to the way we make epistemology. Epistemologists develop models. Our epistemic lives are highly complex, and involve tons of variables. Epistemologists develop simplified models to shed light on this complexity. They enable us to draw some conclusions regarding what is optimal, preferable or conducive of epistemic value. Just think of the following claims:

1. Some combinations of attitudes A and B are dominated.[6]
2. Deferring to expert testimony T maximises expected accuracy.[7]
3. If getting evidence E has a negligible cost, then rational agents pursue E.[8]

But epistemologists' models are similar to those of scientists. Just as scientists sometimes postulate that the ocean floor has infinite depth, we postulate that agents have unbounded cognitive capacities. We assume, for example, that agents have unlimited time and resources for processing the evidence, reason perfectly, or are luminous. This generates skepticism about the relevance of these models. One can wonder whether the above claims are mere artifacts of the model's false assumptions. Perhaps what appears to be

---

2 Levins's (1966: 423) original argument for robustness supports this.

3 See Weisberg (2007). See also Stegenga and Menon (2017).

4 See Odenbaugh (2011) and Fuller and Schulz (2021). See also Eronen (2015).

5 See the dispute between Odenbaugh and Alexandra (2011) and Kuorikoski et al. (2012).

6 See Joyce (1998) and Pettigrew (2016) on dominance reasoning in epistemology.

7 See Titelbaum (2022: chap. 5) on expert principles.

8 See Good (1967).

epistemically preferable, desirable or rational is a mere consequence of the model's inaccurate background assumptions.

One way we can overcome this skepticism is by testing the robustness of the conclusions of idealised models. We could check whether our various epistemic claims are also a consequence of models that make more minimal assumptions concerning the cognitive capacities of agents. Consider the claim that, if getting evidence E is not costly, then rational agents pursue E. With models of bounded rationality, we can try to confirm whether this result is also plausible for agents who have limited cognitive capacities.[9] If we can't reach this conclusion with models of bounded rationality, then the claim is not robust.

This suggests that the robustness of the conclusions derived from ideal epistemological models depends on their confirmation with models of bounded rationality. In other words, if we care about robustness, idealised epistemology can't be the only game in town.[10]

Here is an objection against my argument. I suggest that ideal epistemology consists in the elaboration of simplified models—that is, models of actual agents that make simplifying assumptions. But one could reject this assumption. For instance, perhaps ideals describe hypothetical agents, and we use these descriptions to assess the attitudes of actual agents in terms of their resemblance to the ones ideal agents have. Call this the Resemblance Approach. I don't deny that ideals can simply consist in the description of some hypothetical agents. However, depending on what we want to *learn* from the Resemblance Approach, we need robustness analysis. Suppose, for example, that ideal agents always seek more free information, because this is epistemically beneficial to them. We can, of course, ask ourselves whether actual agents like us also adopt this practice. But if we want to go further, and use this assessment to *evaluate* the practices of current agents, we must also determine whether this practice is relevant, beneficial or appropriate for actual agents. Basically, we need to confirm that the practice's value is not an artifact of the idealising assumptions. In other words, whether the information we

9 As in Hertwig and Engel (2021).

10 Daniel Greco (2023) has argued that some types of idealisations are unavoidable in epistemology. My view does not conflict with his. I do not deny that, in order to elaborate models, we need *some* idealisations regarding agents or their environment. Models are not fully accurate descriptions of the world. However, if we want to assess the robustness of the models in epistemology, we need different ones based on independent assumptions. This includes assumptions about the cognitive capacities of agents. Since we have models that assume that agents have infinite cognitive capacities, we need models of bounded rationality.

gather from the Resemblance Approach is useful depends on some robustness conditions.

3 A Different Type of Robustness and the Exclusivity Argument

There is, however, a similar notion of robustness that doesn't seem to support non-ideal epistemology. This is the notion of robustness found in Carr's (2022) Exclusivity Argument. In this section, I first present this notion of robustness and the argument. Then, I raise concerns regarding the Exclusivity Argument, and I turn the challenge on its head—that is, I argue that this other notion of robustness also supports non-ideal epistemology.

For Carr, the hallmark of non-ideal epistemology is a commitment to the 'ought implies can' principle. For instance, non-ideal epistemologists are trying to accommodate our cognitive imperfections or limitations. As Carr says, 'non-ideal epistemologists hold that what we epistemically ought to do is somehow or other sensitive to our cognitive limitations. And these sorts of sensitivity can be fleshed out as entailing corresponding OIC principles' (Carr 2022: 1137).

Carr thinks there is another important distinction between ideal and non-ideal epistemic norms. She thinks that the former are normatively robust, but not the latter. By normatively robust epistemic norms, she means epistemic norms that 'are both non-conventional and not seriously context-sensitive' (Carr 2022: 1135).[11] In other words, norms are robust when they are not the consequence of contingent factors. In order to figure out whether some epistemic norms are normatively robust, we need to figure out two things:

1. *Is the epistemic norm conventional*? An epistemic norm is not conventional when it is not based on some conventions. And conventions are 'regularities in behaviour that serve some coordinative function, where some alternative regularity in behaviour would have served the same coordinative function equally well, if widely adopted' (Carr 2022: 1135).
2. *Is the epistemic norm seriously context-sensitive*? An epistemic norm is seriously sensitive to context when there is 'no normatively privileged resolution of one or more of the context-sensitive parameters for evaluations of that kind' (Carr 2022: 1135). In other words, if an epistemic norm is seriously sensitive to context, then many contextual factors matter for the evaluation of what we ought to do (or are permitted to do).

Norms that are not normatively robust depend on contingent (historical or personal factors). Conventional norms, for instance, depend on historical developments in communities for regulating the behaviour of epistemic agents. Serious sensitivity to context also depends on contingent factors. For Carr, non-ideal epistemologists who endorse norms that are seriously sensitive to context

11Though technically, the proposal could be that non-ideal epistemic norms are *either* conventional *or* seriously sensitive to context.

make arbitrary choices concerning what counts as an acceptable constraint (Carr 2022: 1152-53). She suggests that non-ideal epistemologists have to draw an arbitrary line between legitimate and illegitimate constraints. This arbitrary line will be selected in accordance with historically or personally contingent factors.

So, there are some putative distinctive properties of ideal epistemic norms. And they give rise to an argument concerning what can be done exclusively by idealised epistemic norms. I call this the Exclusivity Argument.

The argument roughly goes as follows. Carr argues that, if we want to figure out the fundamental nature of epistemic rationality, we need epistemic norms that are not based on contingent factors (what agents like us happen to be capable of doing, historical conventions, or sensitivity to context). Normatively fundamental properties are not dependent on contingent factors. So, if we want to grasp the normatively fundamental properties of epistemology, we need epistemic norms that are normatively robust. And only idealised norms fit this bill. As she says:

> If you go in for a certain kind of metaphysics, then you can think of normatively robust evaluations as latching onto normatively fundamental properties.… Conventions, by contrast, depend on historical contingencies that are independent of these normative joints (Carr 2022: 1136).

Table 1 summarises the differences between ideal and non-ideal epistemology underlying the Exclusivity Argument.

Table 1: Some Putative Differences Between Ideal and Non-Ideal Epistemology

Epistemic norms / claims about these norms	Are these norms sensitive to what agents can(’t) do?	Are these norms conventional?	Are these norms seriously sensitive to context?
Ideal Epistemology (e.g., Bayesianism)	No	No	No
Non-Ideal Epistemology (e.g., bounded rationality)	Yes	Yes	Yes

In the Exclusivity Argument, the notion of robustness at stake is different from the one discussed in §2. However, the two notions are somehow related. Carr is interested in the sort of epistemology that does not depend on *contingent factors*, while philosophers of science are interested in results that do not depend on *false hypotheses*. Both approaches are motivated by a quest for

generality. We are not interested in predictions and conclusions that merely make sense under a narrow class of hypotheses. Yet, the strategy she proposes for getting robust conclusions is completely different. For her, in order to establish robust results, we need to abstract away from contingent factors. In light of the above, we must consider the possibility that, in epistemology, a specific type of robustness is achieved not by the convergence of models, but rather by the analysis of highly idealised models.

4 Bayesian Epistemology and the Availability of Superior Epistemic Norms

The Exclusivity Argument raises questions. It rests on the contentious distinctive properties of epistemic ideals summarised in Table 1. In sections 3 and 4, I take a closer look at these putative distinctive properties.

According to the Exclusivity Argument, ideal epistemology is not concerned with contingent factors, including what agents like us are able to do. In other words, epistemic ideals are insensitive to what agents can or can't do. Most epistemologists, including Carr, think that the Bayesian frameworks that ignore an agent's limited cognitive capacities are idealised (Carr 2022: 1132, 1142). However, it's unclear whether this brand of Bayesian epistemology *completely* ignores what agents can and can't do.

How do we know if a theory or a model is sensitive to what agents can do? Here is a thought: When a theory or a model is sensitive to what agents can do, some better options are excluded from the choices agents have on the ground that these options are somehow unavailable. An option is available just in case there exists a method at our disposal to get the option. Just think of the classical trolley problem (Thompson 1976). A runaway trolley is approaching on the rails next to you. If you do nothing, five workers will perish. If you pull the lever next to you, only one worker will perish. Or course, if you were Superman (or if you could instantly become Superman), you could save everyone on the rails. This would be better than pulling the lever. Unfortunately, you are not Superman, and so this option is unavailable to you. Here, the options you have are sensitive to what you can or can't do. In theory, there are better options than pulling the lever, but these options are somehow unavailable. Or think of the ethics of healthcare rationing (Bognar and Hirose 2014). Policy advisors have to choose how they will allocate the resources available to them on healthcare. They want to make the best use of their limited resources, which often leads to justifiable rationing. Of course, if they had *unlimited* resources, they would make better decisions. So in theory, there are better options than rationing healthcare. But

policy advisors don't have unlimited resources. Accordingly, some options are somehow unavailable to them. This isn't to say that people working on trolley problems and healthcare rationing are unconditional fans of 'ought implies can' (OIC), but some of their arguments factor what agents can and can't do.

Idealised theories of rationality, like Bayesianism, exclude certain options on the grounds that they are somehow unavailable. For instance, idealised theories of rationality say that agents should be responsive to their evidence. But there are better options than that. Just think of the following epistemic norm:

Ignore Misleading Evidence Agents may, or should, form and revise attitudes that are supported by their *non-misleading* evidence, and ignore their misleading evidence.[12]

Strictly speaking, Ignore Misleading Evidence is better than being responsive to the evidence agents have. After all, responding to misleading evidence will lead agents further away from the truth.

Here is another epistemic norm that is superior to forming attitudes in accordance with our total evidence:

Permissive Accuracy Norm Agents may have doxastic attitude D towards P if and only if doxastic attitude D towards P is maximally accurate.[13]

Satisfying the Permissive Accuracy Norm is better than being responsive to the evidence. Responding to misleading evidence will lead agents further away from the truth. By way of contrast, satisfying the Permissive Accuracy Norm can't lead agents away from the truth.

In response to these observations, the ideal rationality theorist could say: 'Agents are not always in a position to discriminate misleading from non-misleading evidence, or to figure out what's accurate. There is no effective and decidable method for satisfying Ignore Misleading Evidence or the Permissive Accuracy Norm. So, these options are somehow unavailable to agents. Theories like Bayesianism describe the best *available* policy agents have' (see, for instance, Horowitz 2014). But this reasoning relies on what agents can and can't do. Availability has to do with the norms that agents *can* operationalise.

12 See Daoust (2021b), Horowitz (2014), Smithies (2016), and Sorensen (2021) for discussion of norms that allow agents to ignore their misleading evidence.

13 See Whiting (2010) for discussion of similar putative requirements. A logically stronger norm of knowledge says this: 'Believe P only if you are in a position to know that P'. An even stronger norm of truth says this: 'Believe P if and only if P'. Carr (2022: 1137) gives two putative reasons why we should not accept the norm of truth, but she does engage with the norm of knowledge, the Permissive Accuracy Norm, and Ignore Misleading Evidence.

This is an indirect way to talk about what agents can do. If we really don't care about what agents can and can't do, then the unavailability of certain norms should be irrelevant.

To make a long story short, it seems that the diagnosis of epistemic ideals underlying the Exclusivity Argument doesn't generalise well. The Bayesian frameworks that ignore an agent's limited cognitive capacities are idealised. Yet, these frameworks are sensitive to what agents are able to do. That is, for Bayesians, agents are not required to satisfy norms that are *unavailable* to them, and availability has to do with what agents *can* operationalise. This is in tension with Carr's claim that idealised theories of rationality are insensitive to what agents can and can't do. At least, this reveals that even idealised theories do not completely ignore what agents can do.

One could raise the following point: Although Bayesian approaches are sensitive to a thin version of OIC, these approaches are not sensitive to stronger interpretations of OIC. For example, Roger White has argued that Bayesian agents must conform to certain general rules, like the Principle of Indifference, even if they are not in a position to know what specific credences they must form in order to comply with them.[14] In other words, we shouldn't expect Bayesian frameworks to offer a fully operational interpretation of every single rule they include. In line with the above, a proponent of Bayesian approaches might qualify Carr's proposal as follows: Ideal theories are insensitive to *some* versions of OIC than non-ideal theories, which is, in part, why they are robust.

For this answer to carry weight, it would be necessary to clarify which interpretation of OIC is fitting of idealisations, and which ones are not. In other words, we would need to introduce a distinction between interpretations of OIC that are adequate for ideal theories, and those that are not. Moreover, these distinctions between different interpretations of OIC cannot be *arbitrary*. If we draw an arbitrary line between interpretations of OIC, we introduce contingent and arbitrary factors into the theory, which conflicts with the aim of robustness described by Carr. The problem is that this sort of neat, unarbitrary classification between variants of OIC is unheard of. Those who want to go down that road owe us an explanation of why the sort of OIC principle I describe is adequate for an ideal theory, while others aren't.[15]

---

14 See White (2010), especially the remarks on pages 167-68.

15 It should be made clear here that I am not equivocating on the term 'can'. I am not, e.g., moving from the logical 'can' to the 'can' of abilities. Throughout my argument, I use an interpretation of 'can' that relates to the uncertainty Bayesian agents have to manage. Bayesian agents have to deal with their own epistemic limits, and they can't know for sure how to satisfy three requirements, namely: Ignore Misleading Evidence,

Without that sort of explanation, a couple of points remain. First, some versions of OIC manifest themselves in idealised models. Second, a dichotomous classification of epistemic theories based on OIC is problematic. And accordingly, arguments based on such a classification are questionable.

Here is another objection to my argument. One could suggest that Bayesian epistemology factors what agents can operationalise, but only because this ideal is relative to a particular domain. There is a domain called rationality (or a sense of 'obligation' that has to do with rationality). Many philosophers think that ideal rationality is 'internal'—namely, that it supervenes on an agent's mental states and processes. Norms like Ignore Misleading Evidence are not internal. They can't be operationalised with a set of rules that are purely internal. And so, norms like Ignore Misleading Evidence are not ideals *of rationality*.[16]

However, if we tie ideals to particular domains, there is no longer a clear distinction between ideal and non-ideal epistemology. Bounded rationality is ideal *relative to certain domains*. For example, one could say that bounded rationality is the best policy for *managing our limited cognitive capacities*. Or, one could say that bounded rationality is the best policy *for dealing with our cognitive imperfections*. So, we can say that bounded rationality is a domain-relative ideal. Domains can be an indirect way to specify some constraints that we want to take into account.

For concreteness, here is an example. Some heuristics allow us to make excellent decisions with very limited cognitive capacities (Todd and Gigerenzer 2000). Does satisfying these heuristics constitute an ideal? Relative to some domains, the satisfaction of fast and frugal heuristics can very well be the best policy we have. For instance, we could say that fast and frugal heuristics are ideal relative to the domain of procedural rationality. Procedural rationality is sensitive to the cognitive limitations agents face (Simon 1976). Just as Bayesian epistemology can be ideal relative to a restricted domain, fast and frugal heuristics can be ideal relative to another restricted domain.

---

the Permissive Accuracy Norm and the Principle of Indifference. It's their inability to overcome uncertainty that compromises the fulfillment of the three preceding norms.

16 See Carr (2022: 1137) on a similar point.

5 Conventional Epistemic Ideals and Nonconventional Bounded Epistemology

5.1 Ideal Conventions

Another worry for the Exclusivity Argument concerns the relationship between ideal epistemic norms and conventions. The Exclusivity Argument rests on the idea that certain historical contingencies, like conventions, are irrelevant to idealised epistemic norms. However, some putative epistemic ideals discussed in print rely on conventions.

First, note that conventions can have many of the usual hallmarks of ideals. Perhaps the clearest illustration of this is the fact that the best solution to many practical problems is, in part, a matter of conventional coordination among agents. Coordination refers to the process of organising individual actions and intentions, so that collectives work together. Road regulations are a good example of this. There is no major advantage to driving on the right rather than driving on the left. What is most important is that everyone drives on the same side of the road. In other words, one of the most important factors for driver safety is coordination among drivers. Other simple problems in decision theory are also about conventions between players. Just think of coordination games in decision theory. Imagine a two-player game with the following payoffs:

		Player 2	
		Option C	Option D
Player 1	Option A	10, 10	0, 0
	Option B	0, 0	10, 10

Here, the ideal solution to the game (that is, the solution that maximises both the expected and actual payoff) is a matter of coordination. If player 1 takes option A, then player 2 should take option C, but if player 1 takes option B, player 2 should take option D. If players can communicate with each other, payoff maximisation is purely a matter of the convention they will go for.

Of course, this is not the only hallmark that conventions and ideals can share. Conventions can be the end-state we want to achieve. They can be the necessary means for full compliance with desiderata that we care about. And their implementation can be highly unrealistic.

Some idealised solutions to problems in epistemology are (in part) about coordination between agents. Think of Kitcher's account of the division of epistemic labour. Kitcher (1990) invites us to consider the following possibility: Suppose you are a 'philosopher-dictator' in charge of designing the perfect

epistemic community. You can design any epistemic community you want (in accordance with properly epistemic goals). Which decisions should you make? Will you ask different researchers to work with different methods? Kitcher argues that, if you were concerned with the advancement of scientific knowledge, you would promote diversity in the epistemic community. Specifically, you would ask different research teams to work with different methods, even under the assumption that one method is more likely than the other to be successful.

The organisation of the scientific community would partly be conventional: What matters to the philosopher-dictator is that some researchers take method A and other researchers take method B, but there are many different ways to coordinate this. The dictator could ask team α to take method A and team β to take method B, but she could ask team β to take method A and team α to take method B. This fits the description of conventional epistemic norms in the Exclusivity Argument. Recall that epistemic norms are conventional when they rely on 'regularities in behaviour that serve some coordinative function, where some alternative regularity in behaviour would have served the same coordinative function equally well, if widely adopted' (Carr 2022: 1135).

According to the Exclusivity Argument, Kitcher's theory of the division of epistemic labour is non-ideal. This is so, because Kitcher's proposal is partly conventional. But recall that, in the Exclusivity Argument, non-ideal epistemic norms have other characteristics. These norms embrace the ought-implies-can principle. They can be implemented in most circumstances by ordinary agents like us. The problem for the Exclusivity Argument is that most agents (or communities) won't be able to implement Kitcher's proposal in most circumstances. The model is not particularly sensitive to what agents can and can't do. Of course, Kitcher doesn't ignore *all* the constraints we face. For instance, his model is sensitive to the resources that are available to epistemic communities. But some of his assumptions are unrealistic.

Here is why. Kitcher begins with the assumption that a philosopher-dictator will design the perfect epistemic community at will. This dictator has control over everything in the community, including the available resources and the workforce. For Kitcher, this assumption is a 'fiction' (Kitcher 1990: 9). This degree of control over the decisions of the scientific community is unlikely to happen. Kitcher also assumes that the probability of success of each method is identifiable by members of the epistemic community. Specifically, he makes 'the supposition that scientists identify the [objective degree of confirmation] that methods will succeed, given an assignment of a number of workers' (Kitcher

1990: 18). Kitcher rightly says that this assumption is an ‘idealisation’. Ordinary scientists are not in a position to identify the objective degree of confirmation that certain methods will succeed in solving problems (conditional on the resources available for testing the method). In view of the foregoing, it’s implausible that Kitcher is describing a non-ideal epistemic norm.

Kitcher’s observations also lead him to think that, in some situations, you can’t meet the demands of individual and collective rationality simultaneously. Suppose that method B is more promising than method A. Scientists *qua individuals* want to take the most promising methods. So, they want to take method B. But scientists *qua members of a group* might sometimes need to take methods that are suboptimal. So, they might need to take method A. The upshot is this: Epistemically speaking, there are situations in which you can’t fully comply with the demands of individual and collective rationality simultaneously. As Kitcher says, there can be an unsolvable ‘mismatch between the demands of individual rationality and those of collective (or community) rationality’ (Kitcher 1990: 6). This is another indication that Kitcher is not committed to the sort of norms that most agents can implement in most circumstances. If he were, he wouldn’t come to the conclusion that agents are sometimes doomed to make irrational decisions, no matter what they do.

So, the theory of the division of epistemic labour causes trouble for the Exclusivity Argument. The theory of the division of epistemic labour is insensitive to what agents can do in different ways. In accordance with the divide between ideal and non-ideal epistemic norms that underlies the Exclusivity Argument, Kitcher’s account of the division of epistemic labour would be an ideal. Yet this theory is conventional. There are many different, conventional ways to implement diversity in the scientific community. That is, the implementation of diversity in a scientific community is, in part, a coordination problem. This is a problem for the Exclusivity Argument, since this view presupposes that ideal epistemic norms abstract away from conventions.

5.2 Non-Ideal Epistemology without Conventions

Relatedly, the Exclusivity Argument says that non-ideal epistemology relies on conventions. However, this claim doesn’t generalise well to a common brand of non-ideal epistemology—namely, constrained expected accuracy optimisation.

Constrained expected accuracy optimisation consists in identifying the best attitudes one can have (in terms of expected accuracy) relative to certain constraints. These constraints can, among other things, reflect what agents can and can’t do. This formal framework is sometimes used for representing the

decisions of non-ideally rational agents.[17] Suppose, for instance, that you know that P's objective probability is 0.5. If you were ideal, you would have a credence of 0.5 in P and a credence of 0.5 in ~P. These credence assignments maximise expected accuracy. However, suppose that you face a constraint: Your credence in P must be greater than or equal to 0.8. So, you are not in a position to maximise expected accuracy. Yet, you could maximise *constrained* expected accuracy (in other words, you can take the attitudes that maximise expected accuracy *relative to the constraints you face*). Here, constrained expected accuracy optimisation would recommend taking a credence of 0.8 in P and a credence of 0.5 in ~P.[18]

The decisions resulting from constrained optimisation have the hallmarks of non-ideal rationality. In the above example, constrained expected accuracy leads agents to violate the Principal Principle, an idealised requirement of rationality. So, the decisions resulting from constrained optimisation might be less than ideal (except, of course, if the set of constraints is empty—in that case, constrained expected accuracy optimisation just is expected accuracy optimisation). At least, since people like Carr think that the standard Bayesian norms are requirements of ideal rationality, it's safe to claim that they would take many decisions resulting from constrained optimisation as non-ideal.

Yet, constrained expected accuracy optimisation is not conventional. Recall that, in the Exclusivity Argument, one hallmark of non-ideal epistemology is that it relies on regularities in the actions and decisions of members of a group. These regularities help members of a group coordinate with each other.[19] This is what Carr means by 'conventions'. But in the above example, your taking a credence of 0.8 in P and a credence of 0.5 in ~P does not depend on the regularities in the behaviour of agents. Constrained expected accuracy optimisation is tangential to whether members of a group coordinate their

17 See, e.g., Arrow (2004) on Simon (1976). See Staffel (2020: 112-16) on constrained optimisation and the approximation of epistemic ideals. Of course, not everyone is on board with representing the decisions of non-ideal agents with constrained optimisation. See, for instance, Todd and Gigerenzer (2000: §2.2) for critical discussion. And constrained optimisation is sometimes used to define ideals—see Wiens (2020) and Lipsey and Lancaster (1956) on operational constraints, optimisation, and first-best scenarios. Yet, it is clear that optimising expected accuracy under constraints would depart from the sort of results found in Bayesian epistemology.

18 Relative to the Brier score, the constrained expected inaccuracy function for this situation goes as follows:

$$Constrained\,expected\,inaccuracy = 0.5\cdot(1-a)^2+0.5\cdot a^2+0.5\cdot(1-b)^2+0.5\cdot b^2, a\geq 0.8.$$

Here, a denotes the agent's credence in P, and b denotes the agent's credence in ~P. We can find the find the credence function that optimises expected accuracy by minimising the above equation under the constraint $a\geq 0.8$.

19 See Dogramaci and Horowitz (2016) and Dogramaci (2015).

behaviour. There can be total anarchy in the epistemic community and it would still be optimal, for you, to take a credence of 0.8 in P and a credence of 0.5 in ~P. These are unrelated issues.

One could object that the operative conventions in a community are sources of constraints. For instance, if the members of a community usually use method M for processing the evidence, then that convention is a constraint restricting how agents can reason. This lends support to the claim that constrained optimisation is conventional: We try to maximise some functions under some constraints, but some constraints come from conventions.[20] However, constrained optimisation doesn't tell us that we have to stick to existing conventions. Suppose again that, in a given community, the convention is to use method M for processing the evidence. However, let's assume that agents are free to follow or ignore the convention. Constrained optimisation does not preclude using a different method for processing the evidence. If all that matters is optimising expected accuracy, perhaps there will be situations in which agents should not stick to method M.

Thus, the claim that non-ideal epistemology is conventional does not generalise well to constrained expected accuracy optimisation.

6 Turning the Challenge on its Head

In this section, I argue that, if we take normative robustness seriously, then we should favour the sort of frameworks that are found in non-ideal epistemology. First, I argue that constrained optimisation reveals an ambiguity in the way we can interpret sensitivity to context. This leads me to conclude that normative robustness supports non-ideal epistemology.

6.1 An Ambiguity Concerning Sensitivity to Context

In the Exclusivity Argument, epistemic norms are seriously sensitive to context when there is 'no normatively privileged resolution of one or more of the context-sensitive parameters for evaluations of that kind' (Carr 2022: 1135). This is supposed to lead to a problem for non-ideal epistemologists. Serious sensitivity to context forces non-ideal epistemologists to make choices concerning what counts as an acceptable constraint (Carr 2022: 1152-53). These choices will be made in accordance with historically or personally

20 Here I ignore the case in which conventions can't be changed or ignored. In these situations, conventions can be constraints. However, we can explain these cases as instances of OIC. That is, in the case of an unchangeable convention X, we merely need to say that agents are not in a position to deviate from X, which is why it's a constraint. The fact that it is a convention is irrelevant.

contingent factors. However, epistemic norms that are selected on the basis of contingent factors won't capture the fundamental features of epistemic normativity.

There is a limited sense in which constrained optimisation is sensitive to context. Agents face different constraints in different contexts. Constrained optimisation doesn't provide a privileged resolution of these context-sensitive parameters. It just says: 'If you face constraint X, take the best option that satisfies X.' So, constrained optimisation is sensitive to context in this specific sense.

But there is also a sense in which it is not. Constrained optimisation is a general principle. What I mean by 'general' is that this principle applies to a wide variety of problems, and is not limited to a specific, contingent class of issues. The selection of the principle 'maximise expected value under constraints' is not made in accordance with historically or personally contingent factors. It does not require that we make specific choices concerning acceptable constraints. The general framework makes sense regardless of the specific context agents find themselves in.

This second way to interpret sensitivity to context matters. A general non-ideal principle that says 'maximise expected value under constraints' does not face the worry described in the Exclusivity Argument. Recall that, according to the Exclusivity Argument, non-ideal epistemologists try to accommodate contingent factors. For instance, the reason why some theorists say that we should reason with heuristics is that they try to accommodate specific contingent factors like 'agents have limited cognitive capacities'. Perhaps the point makes sense for concrete non-ideal norms like 'reason with heuristics', but the point doesn't apply to general principles of constrained optimisation. The principle of maximisation under constraints is not motivated by contingent factors. It is a general framework that can tell us what to do, regardless of the specific context we are in. In other words, the reason why we accept this principle is not that we are trying to accommodate some specific contextual factors.

Principles like 'maximise expected value under constraints' are so general that they do not even preclude the idealised ones. For instance, expected accuracy maximisation can be subsumed under constrained expected accuracy maximisation. That is, unconstrained optimisation is constrained optimisation, but relative to an empty set of constraints. As such, constrained optimisation is

general enough to encompass the sort of project that many ideal epistemologists are pursuing.[21]

For concreteness, here is an example illustrating this. How can we justify the Principal Principle? One justification of the Principal Principle relies on unconstrained optimisation. It says that satisfying the Principal Principle maximises expected accuracy (Pettigrew 2016). In a constrained optimisation framework, this justification of the Principal Principle is a particular instance of the general framework. That is, in cases where agents face an *empty set of constraints*, the Principal Principle maximises expected accuracy. It's just that there are other contexts in which other decisions are optimal. The point is that constrained accuracy optimisation encompasses the sort of project that ideal epistemologists are interested in. It just covers more cases.

6.2 Non-Ideal Principles and the Foundations of Epistemic Normativity

The Exclusivity Argument rests on putative distinctive properties of epistemic ideals. So far, I have presented four observations that are in tension with the putative distinction between ideal and non-ideal epistemic norms. Table 2 summarises them.

21 It is not uncommon that a general principle finds different interpretations in different contexts. Bayesians would agree with that. Take the Principal Principle. Suppose that I know that P's objective probability is 0.5, but you don't know that. If we are ideally rational, we might end up with different credences in P that both satisfy the Principal Principle. The reason why we might not end up with the same credence in P is contextual: We do not have the same knowledge.

Table 2: Some Problematic Observations Concerning Ideal and Non-Ideal Epistemic Norms

Epistemic norms / claims about these norms	Are these norms sensitive to what agents can(’t) do?	Are these norms conventional?	Are these norms seriously sensitive to context?
Bayesianism (Ideal)	Partly: Bayesian epistemology is sensitive to what agents can operationalise	No	No
Division of Epistemic Labour (Ideal)	Mostly not	Yes	Undetermined
Constrained Optimisation (Non-Ideal)	Yes	No	Ambiguous: The general principle ‘maximise value under constraints’ makes sense independently of contingent, contextual factors.

In addition to casting doubt on the Exclusivity Argument, the points made in the previous sections pave the way for a positive answer to the initial question Carr is interested in. We are trying to identify a type of epistemology that can capture the fundamental nature of epistemic normativity. Figuring out the foundations of epistemic normativity is an important project that has drawn a lot of attention (see, among others, Côté-Bouchard 2016, Cowie and Greenberg 2018, Floweree 2018, and Greco 2010). Ideal epistemology, Carr says, is the ‘only game in town’ for carrying out this task (Carr 2022: 1131). However, if we take the observations from the previous sections seriously, it appears that non-ideal epistemology has better resources to capture the fundamental nature of epistemic normativity.

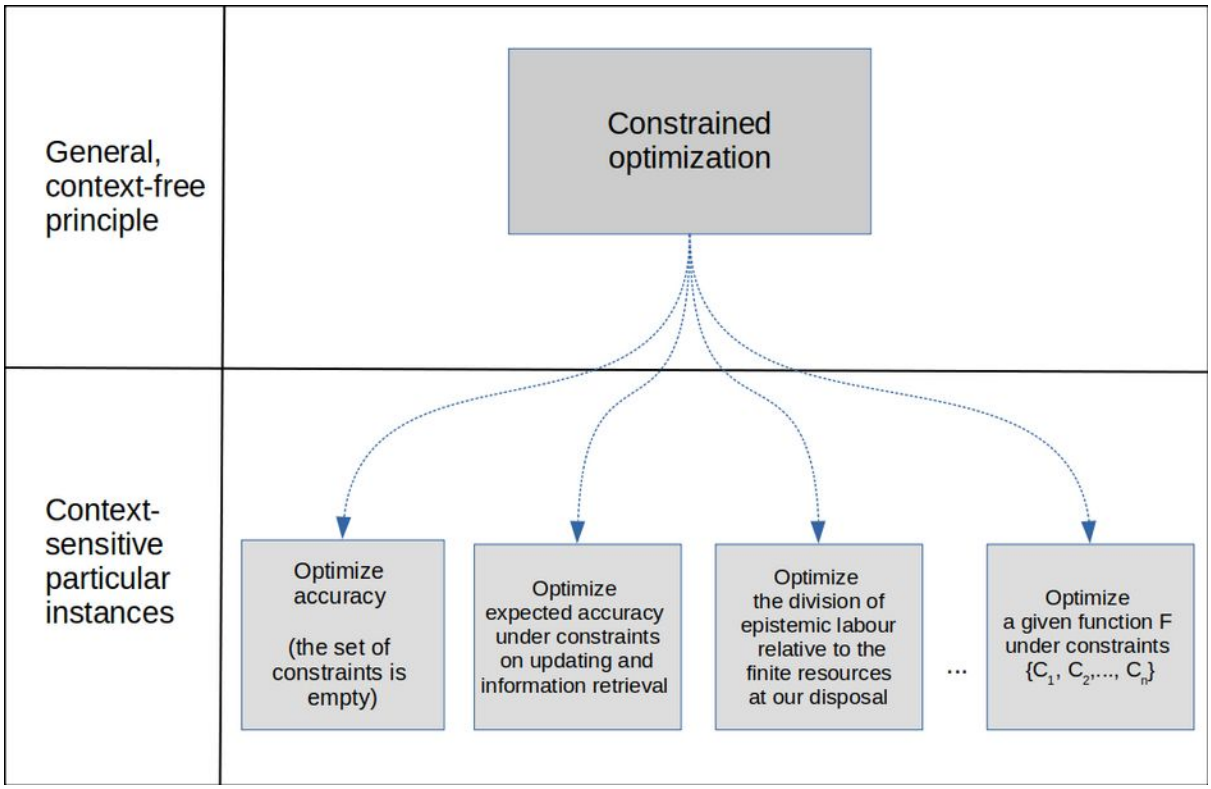
Here is why. A general principle like ‘maximise value under constraints’ can capture the normatively fundamental features of epistemology. As we saw in section 5.1, this general principle is not, in itself, sensitive to context. The particular interpretations of this principle are sensitive to context, but the principle itself isn’t. So, we can capture the fundamental normative features of epistemology with this principle. That is, one could say that the most basic principle of epistemology is to maximise epistemic value under constraints (this

is just an example, of course)[22]. The specific interpretation of this principle might vary from a context to another, but the general principle can hold in every context. This principle does not depend on the acceptance of contingent factors.

In fact, if we want to capture the normatively fundamental features of epistemology, a general principle like ‘maximise value under constraints’ *fares better* than a principle of unconstrained optimisation. As we saw in section 5.1, expected accuracy maximisation is a particular instance of constrained expected accuracy maximisation. Unconstrained maximisation is just constrained maximisation relative to an empty set of constraints. Not facing constraints, or ignoring constraints, is just *one context among others*. If we take the idea that sensitivity to context is an obstacle to capturing the fundamental normative features of epistemology, we should not make particular assumptions about the constraints agents face. This is exactly what the general principle of constrained optimisation does. Figure 1 shows how various epistemic theories can be captured with the general principle of constrained optimisation.

22 Again, some theorists deny that we can represent the decisions of non-ideal agents with constrained optimisation.

Figure 1: Constrained optimisation as a general, context-free principle



Principles that come from non-ideal epistemology, like constrained optimisation, are in a better position to capture the normatively fundamental features of epistemology. If you really care about the normative foundations of epistemology, and wish to avoid contextual hypotheses, you can't say: 'We'll start by assuming that agents have infinite cognitive capacities, reason perfectly and instantly, and so forth'. These are not general hypotheses. At a minimum, these hypotheses are not more general than stipulating that agents have limited cognitive capacities. What you can say, on the other hand, is: 'We'll look for ways to do epistemology that avoid problematic commitments concerning the constraints agents face, their cognitive capacities, and so forth'. And this is exactly what general principles like constrained optimisation give you.

7 Conclusion

There are two reasons for thinking that non-ideal approaches should have a more central place in epistemology. They are closely linked to the question of the robustness of the conclusions we draw from models.

The first reason is that models based on simplified or false assumptions, such as the assumption that the ocean floor has infinite depth, generate skepticism. More specifically, for any conclusion C of the model, one may wonder whether C is not an artifact of the false assumptions on which the model is based. In order

to overcome this skepticism, we should confirm whether C is a consequence of other models, that is, frameworks that make different assumptions. This way, we know that the conclusion is robust, in the sense that it is not sensitive to problematic background assumptions. The same argument applies to epistemology. We have idealised models. They make the assumption that agents have infinite cognitive capacities. These models can generate skepticism. We must also try to determine whether other models—those in which agents have limited resources—allow us to draw the same conclusions. So, we need to do more than ideal epistemology.

The second argument is about the sort of principles that can capture the normatively fundamental features of epistemology. This is what Carr (2022) calls normatively robust epistemology. She argues that idealising is the only way to reach normatively robust results. In other words, she proposes an Exclusivity Argument in favour of ideal epistemology. I have argued that some approaches found in non-ideal epistemology fare better than idealised ones to elucidate the normatively fundamental features of epistemology.

Thus, we have turned the challenge on its head. You want to figure out the normatively fundamental features of epistemology? Go non-ideal.

Acknowledgements

I am very grateful to anonymous referees for detailed, useful feedback on this project. Thanks also to David Thorstad for helpful comments.

Funding Information

This research was supported by the Fonds de recherche du Québec – Société et culture (grant #268137) and the Social Sciences and Humanities Research Council (grant #430-2022-00065).

ORCID

Marc-Kevin Daoust, ORCID: 0000-0002-6710-1944

References

Arrow, Kenneth J. (2004) ‘Is Bounded Rationality Unboundedly Rational? Some Ruminations’, in Mie Augier and James G. March, eds., *Models of a Man: Essays in Memory of Herbert A. Simon*: 47-56. MIT Press.

Bognar, Greg and Iwao Hirose (2014) *The Ethics of Health Care Rationing: an Introduction*. Routledge.

Carr, Jennifer R. (2022) ‘Why Ideal Epistemology?’, *Mind* **131**: 1131–1162

Christensen, David (2007) ‘Does Murphy’s Law Apply in Epistemology? Self–Doubt and Rational Ideals’, in Tamar Szabó Gendler and John Hawthorne, eds., *Oxford Studies in Epistemology 2*: 3–31. Oxford University Press.

Christensen, David (2010) ‘Higher-Order Evidence’, *Philosophy and Phenomenological Research* **81**: 185–215.

Côté-Bouchard, Charles (2016) ‘Can the Aim of Belief Ground Epistemic Normativity?’, *Philosophical Studies* **173**: 3181-3198.

Cowie, Christopher and Alexander Greenberg (2018) ‘Constitutivism About Epistemic Normativity’, in Christos Kiriakou and Robin McKenna, eds., *Metaepistemology. Palgrave Innovations in Philosophy*: 173-196. Palgrave Macmillan.

Daoust, Marc-Kevin (2021a) ‘Adversariality and Ideal Argumentation: A Second-Best Perspective’, *Topoi* **40**: 887-898.

Daoust, Marc-Kevin (2021b) ‘False Beliefs and Misleading Evidence’, *Theoria* **87**: 520-541.

Daoust, Marc-Kevin (forthcoming). ‘What Second-Best Epistemology Could Be’, *Analytic Philosophy*.

DiPaolo, Joshua (2019) ‘Second Best Epistemology: Fallibility and Normativity’, *Philosophical Studies*, **176**: 2043-2066.

Dogramaci, Sinan (2015) ‘Communist Conventions for Deductive Reasoning’, *Noûs* **49**: 776-799.

Dogramaci, Sinan and Sophie Horowitz (2016) ‘An Argument for Uniqueness about Evidential Support’, *Philosophical Issues* **26**: 130-147.

Eronen, Markus (2015) ‘Robustness and Reality’, *Synthese* **192**: 3961-3977.

Flowerree, Amy (2018) ‘Epistemic Schmagency?’, in Christos Kiriakou and Robin McKenna, eds., *Metaepistemology. Palgrave Innovations in Philosophy*: 289-310. Palgrave Macmillan.

Fuller, Gareth P. and Armin W. Schulz (2021) ‘Idealizations and Partitions: A Defense of Robustness Analysis’, *European Journal for Philosophy of Science* **11** (paper 107)*:* 1-15.

Good, Irving J. (1967) ‘On the Principle of Total Evidence’, *British Journal for the Philosophy of Science* **17**: 319-321

Greco, John (2010) *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge University Press.

Greco, Daniel (2023) *Idealization in Epistemology: a Modest Modeling Approach*. Oxford University Press.

Hertwig, Ralph and Christoph Engel, eds. (2021) *Deliberate Ignorance: Choosing not to Know*. MIT Press.

Horowitz, Sophie (2014) ‘Immoderately rational’, *Philosophical Studies* **167**: 41-56.

Joyce, James M. (1998) ‘A Nonpragmatic Vindication of Probabilism’, *Philosophy of science* **65**: 575-603.

Karlan, Brett (2021) ‘Reasoning With Heuristics’, *Ratio* **34**: 100-108.

Kitcher, Philip (1990) ‘The Division of Cognitive Labor’, *The journal of philosophy* **87**: 5-22.

Kuorikoski, Jaakko, Aki Lehtinen, and Caterina Marchionni (2012) ‘Robustness Analysis Disclaimer: Please Read the Manual Before Use!’, *Biology & Philosophy* **27**: 891-902.

Levins, Richard (1966) ‘The Strategy of Model Building in Population Biology’, *American scientist* **54**: 421-431.

Lipsey, R. G. and Kevin Lancaster (1956) ‘The General Theory of Second Best’, *The Review of Economic Studies* **24**: 11-32.

McKenna, Robin (2023) *Non-Ideal Epistemology*. Oxford University Press

Odenbaugh, Jay (2011) ‘True Lies: Realism, Robustness, and Models’, *Philosophy of Science* **78**: 1177-1188.

Odenbaugh, Jay and Anna Alexandrova (2011) ‘Buyer Beware: Robustness Analyses in Economics and Biology’, *Biology & Philosophy* **26**: 757-771.

Pettigrew, Richard (2016) *Accuracy and the Laws of Credence*. Oxford University Press.

Sampson, Eric (2022) ‘What if Ideal Advice Conflicts? A Dilemma for Idealizing Accounts of Normative Practical Reasons’, *Philosophical Studies* **179**: 1091–1111

Simon, Herbert A. (1976) ‘From Substantive to Procedural Rationality’, in T.J. Kastelein et al., eds., *25 Years of Economic Theory*: 65-86. Springer.

Smith, Michael (1994) *The Moral Problem*. Blackwell.

Smithies, Declan (2016) ‘Belief and Self-Knowledge: Lessons from Moore’s Paradox’, *Philosophical Issues* **26**: 393-421.

Sorensen, Roy (2021) ‘Stealing Harman’s Thought: Knowledge Saboteurs and Dogmatists’, *Synthese* **198**: 1787-1799.

Staffel, Julia (2020) *Unsettled thoughts: A theory of degrees of rationality*. Oxford University Press.

Stegenga, Jacob and Tarun Menon (2017) ‘Robustness and Independent Evidence’, *Philosophy of Science* **84**: 414-435.

Thomson, Judith Jarvis (1976) ‘Killing, Letting Die, and the Trolley Problem’, *The Monist* **59**: 204-217.

Titelbaum, Michael G. (2022) *Fundamentals of Bayesian epistemology 1: Introducing Credences*. Oxford University Press.

Todd, Peter M. And Gerd Gigerenzer (2000) ‘Précis of Simple Heuristics that Make us Smart’, *Behavioral and Brain Sciences* **23**: 727-741.

Valentini, Laura (2012) ‘Ideal vs. Non-Ideal Theory: A Conceptual Map’, *Philosophy Compass* **7**: 654-664.

Weisberg, Michael (2006) ‘Robustness Analysis’, *Philosophy of science* **73**: 730-742.

White, Roger (2010) ‘Evidential Symmetry and Mushy Credence’, in Tamar Szabó Gendler and John Hawthorne, eds., *Oxford studies in epistemology*, *3*: 161-186. Oxford University Press.

Whiting, Daniel (2010) ‘Should I Believe the Truth?’, *Dialectica* **64**: 213-224.

Wiens, David (2020) ‘The General Theory of Second Best is More General than you Think’, *Philosopher’ Imprint* **20**: 1-26. HDL: 2027/spo.3521354.0020.005