

*Idealization in epistemology: A modest modeling approach*, by Daniel Greco, Oxford, UK: Oxford University Press, 2023. Pp. 208.

David Thorstad  
Review forthcoming in *Mind*. Please cite published version.

## 1 Introduction

*Idealization in epistemology* is an ambitious and novel attempt to grapple with the role of idealization in epistemology. It is refreshingly clear, argument-rich, and full of case studies that can be read on their own, as well as part of a broader modeling project. *Idealization in epistemology* is a worthwhile read not only for epistemologists, but also for philosophers of science, political philosophers, and others interested in the role of idealization and modeling in philosophical and scientific practice.

In this review, I outline the project of the book (§2) and extract lessons (§3). I also raise challenges focusing on the privileged place of Bayesian theorizing within Greco's approach (§4) and some anti-realist tendencies (§5) that go beyond the methodological lessons of Greco's modeling project. Section 6 concludes.

## 2 Greco's project

Greco distinguishes two projects that modelers might have. *Ambitious modelers* aspire towards a single, fully accurate theory of a domain. By contrast, *modest modelers* are content with a patchwork of overlapping models, each partial and less-than-fully accurate. Greco suggests that philosophers may have become too taken with disciplines such as fundamental physics in which ambitious modeling is the norm. When we widen our lens to disciplines such as economics, we glimpse the possibility of a mature and explanatorily powerful project of modest modeling.

Greco cashes out the methodological upshots of modest modeling in terms of a distinction due to Mike Titelbaum (2012) between *models*, which are particular representations,

and *modeling frameworks*, which are recipes for building models of a similar kind. Within this framework, modest modeling suggests flexibility in framework selection.

**Modesty in Framework Selection:** A modest modeler may use different frameworks to model different phenomena, relying essentially on good judgment in making her selection, and without any hope of an uber-framework that would provide systematic rules for which framework to use when. (Greco 2023, p. 36)

Modest modeling counsels similar flexibility in creating models within a given modeling framework:

**Modesty in Model Selection:** Once the modest modeler has settled on a framework to use to model some phenomenon, the question of which model within that framework to construct may be similarly unsystematic. Perhaps the different models a framework allows the modeler to construct are in principle incapable of being merged into a single model, and good judgment is required to determine which models to use for which purposes. (Greco 2023, p. 36)

We will see in Section 5 that modest modeling may have metaphysical as well as methodological implications.

The bulk of *Idealization in epistemology* is devoted to a series of case studies. The unifying theme of these case studies is that modest modeling makes room for a range of approaches that have recently come under fire for being, in one or another sense, too idealized. Chapter 3 defends fragmentationist theories of belief as solutions to the problems of coarse grain facing possible-world semantics. Chapter 4 defends the Bayesian assumption that agents are certain of what they learn by noting that even alternatives, such as Jeffrey conditionalization, entail forms of certainty such as the immunity of knowledge to undercutting defeat. Chapter 6 defends level-bridging principles, such as the KK principle that agents know what they know, by bringing them within the fragmentationist

approach articulated in Chapter 3. And Chapter 7 defends the possibility of common knowledge against a recent argument by Harvey Lederman (2018) that margin-of-error principles render common knowledge impossible.

While modest modeling is a pluralistic enterprise, Greco is careful to emphasize that not everything goes for modest modelers. In Chapter 5, Greco argues that modest modelers do not need to make use of full-belief in modeling: probabilistic models of agents' belief states can do everything that is needed. I return to this issue in Section 4.

*Idealization in epistemology* concludes with a series of reflections on the debate between ideal and non-ideal epistemologists. Many of these reflections contain important lessons for epistemology, so a good place to begin engagement with *Idealization in epistemology* will be to set out these and other lessons that epistemologists might learn from Greco.

### **3 Lessons from Greco**

*Idealization in epistemology* contains a number of important lessons for epistemologists. First and foremost, Greco emphasizes that all models are idealized. The mere fact of idealization is no objection to a model. Quite the opposite, it may be a sign that the model is doing its job, and idealization is becoming an indispensable tool in a world increasingly driven by mathematical models.

In this sense, Greco emphasizes, we are all ideal theorists. For this reason, it may not be productive to formulate epistemological debates in terms of a fundamental opposition between ideal and non-ideal theory. In objecting to idealized models, we might do better to point out fundamental features of human cognition that are often left out of these models, such as cognitive processes (Simon 1976), task environments (Todd and Gigerenzer 2012), or limited cognitive abilities and the costs of exercising them (Gigerenzer and Selten 2001). In each case, we must argue explicitly and in detail for the importance of incorporating omitted factors into epistemological models, and we should be open to the response that incorporating these features is unhelpful in the present context.

Furthermore, Greco stresses, the mere fact that a model fits the world imperfectly is no objection to the model. Although Greco would not put the point quite this way, the only perfect model of the world is the world itself, and that is no model at all. If we want to object to a model, we need to provide a better model, one which fits the world more accurately than previous models or exceeds them in important theoretical respects. Otherwise, we may be doing no more than telling the modeler what she already knows: that there are some respects in which her model fits the world imperfectly.

Greco also emphasizes the domain-relativity of modeling aspirations. Philosophers are often quite taken by the ambition of fields such as fundamental physics which seek, on some interpretations, to provide a single correct and totalizing model of all relevant phenomena. Greco rightly stresses that these aspirations are not always achievable or helpful in many domains. In particular, social sciences such as economics are often willing to settle for a patchwork of imperfectly predictive models, without seeking to combine these models into a single totalizing worldview.

Greco does important work in emphasizing how de-idealization efforts can fail to resolve, or even exacerbate the problems they were introduced to solve. For example, Greco notes that on at least one common complexity criterion, the 'level-k' knowledge assumption that common knowledge stops after a finite level of iterations favored by some game theorists, philosophers and linguists actually makes an agent's cognitive state more complex than it would be if common knowledge iterated indefinitely. Likewise, Greco notes that tools such as Jeffrey conditionalization introduced to remove the assumption of certainty still incorporate certainty-like assumptions, such as the impossibility of undercutting defeat.

Finally, Greco pays careful attention to modeling developments in fields such as economics and computer science that are not always incorporated into epistemological debates. In particular, Greco's proposal to analyze common knowledge using distributed systems models and Greco's use of complexity theory to unpack the complexity of iterated belief states are valuable contributions that should be taken seriously in subsequent

debates.

These lessons make *Idealization in epistemology* an important contribution to the literature. There are some points on which further discussion might focus. In the rest of this review, I want to highlight two places where some readers might differ from Greco. In each case, I think that the disagreement is separable from the underlying project: modest modelers have the resources to go either way.

## 4 Bayesian hegemony

Rightly or wrongly, Bayesian theorizing has earned a reputation across many disciplines as a hegemonistic project. Many Bayesians do not merely aim to use Bayesian tools to account for a few descriptive and normative phenomena, but increasingly hold out hope to describe many or all phenomena in Bayesian terms.

Readers might expect Greco's book to stay comfortably away from hegemonistic Bayesian projects. The modest modeler has no objection to multiple models being used for different purposes: indeed, this is exactly what she expects. But at times, Greco is quite strongly set in favor of Bayesian models. For example, Chapter 5, on belief and credence, argues that there is no need to invoke full belief in descriptive or normative modeling: anything full belief can do could be done at least as well by credal models. This is, as Greco notes, a surprising stance for a modest modeler. Chapter 5 concludes with the following observation:

The arguments of this chapter may seem in tension with the modest, pluralistic methodological stance urged throughout this book. Shouldn't a modest modeler be open to the suggestion that we've tried to bring too wide and diverse a range of phenomena within the ambit of decision theory? Shouldn't she celebrate the proliferation of modeling frameworks that might carve off some terrain that ambitious, imperialistic types – economists, especially – have prematurely claimed for decision theory? (Greco 2023, p. 116)

In the case of full belief, Greco goes on to answer these questions in the negative. I think that this stance may be more restrictive than current evidence allows.

Earlier in the same chapter, Greco considers the view that credences may sometimes be too complex for bounded agents to form and use in reasoning, and that reasoning might sometimes be productively simplified through the use of coarse-grained attitudes such as full belief. Greco replies, with little argument, that it should be possible for limited agents to make do with ‘small-world’ Bayesian models, and that these models will be appropriately useful while respecting computational constraints.

While this view is far from refuted, it is surprising enough to deserve more careful defense. Consider, for example, the psychology of reasoning. Until perhaps two decades ago, it was assumed by most leading theories that human reasoning is structured on the model of coarse-grained attitudes manipulated through logical reasoning (Johnson-Laird 1983; Rips 1994), rather than fine-grained attitudes manipulated through probabilistic reasoning. While there has been a resurgence of Bayesian views of reasoning (Oaksford and Chater 2007, 2020), these views are by no means accepted matters of scientific fact, and many of our best descriptive accounts of human reasoning continue to describe reasoning in coarse-grained terms.

One reason why it would be appropriate for modest modelers to engage more fully with existing empirical evidence is that, as suggested above, they could be wrong: it is a live scientific hypothesis that humans have many coarse-grained attitudes. A second reason to engage with these discussions is that the types of fine-grained theories which have garnered the most empirical support are substantially different from the types of Bayesian theories often found in epistemological discussions. Bayesian cognitive scientists now commonly hold that agents make judgments and decisions by randomly sampling portions of their total evidence, rather than surveying all evidence held in memory (Icard 2018); that Bayesian reasoning is sometimes done through heuristics which only approximate the probability calculus (Oaksford and Chater 2007); or that Bayesian processes may be a better description of higher-order reasoning processes than of first-order reasoning

processes (Lieder and Griffiths 2017). While there is certainly room for Bayesian theories of human reasoning, it is not so clear that those theories will be reducible to the kinds of small-world Bayesian models found in traditional discussions. For example, the Bayesian agents just described can hold inconsistent attitudes (Zhu et al. 2020); make logical mistakes (Oaksford and Chater 2007); and show familiar cognitive biases (Dorst forthcoming; Lieder et al. 2018).

A final reason to engage more deeply with these empirical discussions is that the success of Bayesian theories may be, in the spirit of modest modeling, domain-specific. Bayesian accounts are very popular in domains such as perception and memory, but they have been much less successful in domains such as metacognition (Proust 2013) and risk perception (Slovic et al. 2007). Consider metacognition: the processes through which agents monitor and control memory, reasoning and other cognitive processes. The received view of metacognition is that there are two systems of human cognition: a rarely-used belief-based system that works exactly as Bayesians describe, and a commonly-used affective system that is not driven by beliefs of any kind, probabilistic or otherwise (Dunlosky and Metcalfe 2008; Koriat 1997). For example, reasoning may be overseen by metacognitive feelings of confidence, error, or problem solvability that differ importantly from their judgmental counterparts (Ackerman and Thompson 2017).

The turn against Bayesian accounts of metacognition was not driven by theoretical speculation, but rather by empirical evidence. We know, for example, that people monitor their own memory and reasoning differently from how they monitor the memory and reasoning of others (Koriat et al. 2004; Kornell and Bjork 2009). It is very natural to explain these differences by positing an affective system by which our own memory, reasoning and other processes are accompanied by a special class of metacognitive feelings, driven by our own experience of memory and reasoning, which could not be activated in monitoring the memory or reasoning of others, of which we have no experience. By contrast, differences in self-directed and other-directed monitoring are quite puzzling to explain if the same Bayesian processes are activated in both cases.

A modest modeler might object that even if non-Bayesian models are more accurate in some domains, there may be other grounds on which Bayesian models should be preferred. For example, perhaps non-Bayesian models are too complex, or provide only limited descriptive improvement. In some cases, that is exactly the right thing to say. But the past several decades have seen growing recognition in many fields of the fact that non-Bayesian models can be both tractable and predicatively indispensable. Returning, for example, to the field of economics, Richard Thaler's 'Anomalies' column is widely credited with convincing mainstream economists that seemingly non-Bayesian phenomena such as endowment effects (Kahneman et al. 1991) and mental accounting (Thaler 1990) lead to important predictive deviations from classical theories, and behavioral economics models have become tractable and important enough to bring home at least three Nobel Prizes within the field.

If this is right, then we arrive at a tentative view more in the ballpark of what might be expected from modest modelers. Bayesian accounts may be right in some domains, and wrong in others, and in every case it is important to investigate the specific evidence in order to determine in detail the type of Bayesian or non-Bayesian account that might be given. The success of Bayesian accounts is not merely a matter of descriptive accuracy, but also of theoretical virtues such as simplicity. The correct Bayesian account may be a simple 'small world' version of traditional models, but in some domains it may be much more complex than this. All of this is, in itself, no objection to Greco's project, but rather a suggestion of how the project might be elaborated in a more ecumenical direction.

## 5 Anti-realism

As we have seen, Greco takes modest modeling to have two methodological upshots:

**Modesty in Framework Selection:** A modest modeler may use different frameworks to model different phenomena, relying essentially on good judgment in making her selection, and without any hope of an uber-framework that would

provide systematic rules for which framework to use when. (Greco 2023, p. 36)

**Modesty in Model Selection:** Once the modest modeler has settled on a framework to use to model some phenomenon, the question of which model within that framework to construct may be similarly unsystematic. Perhaps the different models a framework allows the modeler to construct are in principle incapable of being merged into a single model, and good judgment is required to determine which models to use for which purposes. (Greco 2023, p. 36)

Modesty in Model and Framework Selection are sensible methodological principles. Modelers often use different models and frameworks for different purposes. These models and frameworks have many different assumptions, and it would be tedious, at best, to chase down the exact relationship between these assumptions.

However, Greco does not just want to make a methodological point about modeling. Greco often adopts an anti-realist approach in which there *is* no complete and fully accurate theory of many of the domains modeled by social scientists, and on which discourse about these domains is justified in part by its usefulness for us. For example, Greco reads Daniel Dennett, David Wallace and Sean Carroll as motivating the view that:

There's a pragmatic element to non-fundamental ontology. The justification for describing reality in fundamental terms - whatever those turn out to be - is quite simple; that's just how the world is. But the justification for describing the world in non-fundamental terms - e.g., in chemical, or computational, or mental terms - is partly self-regarding: such descriptions are humanly tractable. (Greco 2023, p.33)

This type of anti-realist perspective about the objects of scientific modeling should be taken seriously, and it is refreshing to see the claim brought into epistemology from its traditional home in the philosophy of science. While I do not want to re-litigate classic

debates between realists and anti-realists, I do want to highlight two places where Greco's anti-realist commitments may be playing an important role in the manuscript.

First, Greco defends a version of model contextualism on which different models of an agent's epistemic state are appropriate in different modeling contexts. Greco does not mean this in the familiar sense that we may sometimes use different modeling frameworks to describe an agent's belief state. Rather, Greco urges that two outright inconsistent models of the same type may be used to describe the same agent in different contexts.

For example, consider the classic fragmentationist picture of David Lewis (1982), who used to believe that Nassau Street runs east-west, that the railroad nearby runs north-south, and that the two are roughly parallel. On an orthodox view, Lewis has a single mental state broken into two fragments. The first fragment contains the belief that Nassau Street runs east-west, and the second fragment contains the belief that the railroad runs north-south. Greco urges a different view on which Lewis should be modeled as having a single, logically consistent belief state. Which state? Greco proposes that we should change the state we model Lewis as occupying based on our purposes as modelers. To explain cases in which Lewis takes Nassau Street to travel east, we should model Lewis as occupying a superset of the first fragment. To explain cases in which Lewis takes the railroad to travel north, we should model Lewis as occupying a superset of the second fragment. This proposal differs from the orthodox fragmentationist perspective by not modeling Lewis as holding inconsistent beliefs at a single time, and also by deliberately not providing a precise statement of when each fragment is activated, for fear of collapse into a traditional fragmentationist picture.

I think that many readers of a more realist persuasion may find this type of contextualism difficult to swallow. It does not follow from the methodological necessity of using modeling frameworks with different types of assumptions. And if there is, at root, a true fact about what mental state Lewis occupies, we should probably hope as a methodological matter to do better than to settle on a pair of inconsistent descriptions of Lewis' mental state.

A similar issue occurs in the final chapter, where Greco confronts the idea that agents may, as a matter of fact, not have logically or probabilistically consistent beliefs. Inconsistency is generally taken to be an important finding from social scientific research on human judgment and decisionmaking, which does not merely evidence the bare fact of inconsistency but also aims to provide precise descriptions of the patterns of inconsistency that humans exhibit. Greco is not sure that this is correct: while ‘modeling us as having probabilistically coherent credences involves idealization, it doesn’t follow that we *actually* have probabilistically *incoherent* credences” (p. 171).

If we take this statement as a methodological claim about modeling, it contains an important insight: even bounded rationality theorists sometimes model agents as holding coherent credences, in order to facilitate tractable discussion of other features of agents’ cognitive lives. But the metaphysical claim that agents do not, as a matter of fact, have inconsistent credences is quite a bit stronger than this. From a more traditional realist perspective, this claim can look like a way of denying some of the most significant scientific discoveries about human cognition in the past century.

This type of anti-realist stance may also raise problems when we turn from descriptive to normative modeling. Greco quite rightly stresses that models can be useful guides to first-personal deliberation, even if agents do not hold attitudes precisely corresponding to anything in the model. But when we turn from first-personal guidance to third-personal normative evaluation, matters become trickier. For example, many philosophers think it is a requirement of rationality that agents’ attitudes be coherent. While these philosophers may be willing to reformulate their view in terms of beliefs, credences, or other attitudes, they will want to be reassured that there is some way or another to make sense of normative coherence requirements and the application of these requirements as third-personal evaluative standards. It is not clear how this would work if, in fact, there is no fact of the matter about whether an agent’s attitudes are coherent.

That is not to say that Greco has nothing to say for himself in this regard. After all, Greco reminds us, quasi-realists have been in the business of mimicking third-personal

evaluative statements for a very long time, and they have gotten quite good at it. Perhaps some such strategy could be extended to account for epistemic evaluations, such as third-person coherence norms. It would be an interesting project to develop this story further.

## 6 Conclusion

*Idealization in epistemology* is a worthwhile read for philosophers interested in modeling and idealization, as well as for those concerned with the fate of idealized views about common knowledge, level-bridging principles, certainty and coherence. It may also be worthwhile to explore connections between modest modeling and domains not covered in this book, such as idealization in political philosophy.

DAVID THORSTAD

Vanderbilt University, United States

david.thorstad@vanderbilt.edu

## References

Ackerman, Rakefet and Thompson, Valerie. 2017. "Meta-reasoning: Monitoring and control of thinking and reasoning." *Topics in Cognitive Science* 21:607–617.

Dorst, Kevin. forthcoming. "Rational polarization." *Philosophical Review* forthcoming.

Dunlosky, John and Metcalfe, Janet. 2008. *Metacognition*. Sage.

Gigerenzer, Gerd and Selten, Reinhard (eds.). 2001. *Bounded rationality: The adaptive toolbox*. MIT press.

Greco, Daniel. 2023. *Idelization in epistemology: A modest modeling approach*. Oxford University Press.

Icard, Thomas. 2018. "Bayes, bounds, and rational analysis." *Philosophy of Science* 85:79–101.

Johnson-Laird, Philip. 1983. *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

Kahneman, Daniel, Knetsch, Jack, and Thaler, Richard. 1991. "Anomalies: The endowment effect, loss aversion, and status quo bias." *Journal of Economic Perspectives* 5:193–206.

Koriat, Asher. 1997. "Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning." *Journal of Experimental Psychology: General* 126:349–70.

Koriat, Asher, Bjork, Robert, Sheffer, Limor, and Bar, Sarah. 2004. "Predicting one's own forgetting: The role of experience-based and theory-based processes." *Journal of Experimental Psychology: General* 133:643–56.

Kornell, Nate and Bjork, Robert. 2009. "A stability bias in human memory: Overestimating remembering and underestimating learning." *Journal of Experimental Psychology: General* 138:449–68.

Lederman, Harvey. 2018. "Uncommon knowledge." *Mind* 1069–1105.

Lewis, David. 1982. "Logic for equivocators." *Noûs* 16:431–41.

Lieder, Falk and Griffiths, Thomas. 2017. "Strategy selection as rational metareasoning." *Psychological Review* 124:762–94.

Lieder, Falk, Griffiths, Thomas, Huys, Quentin, and Goodman, Noah. 2018. "The anchoring bias reflects rational use of cognitive resources." *Psychonomic Bulletin and Review* 25:322–49.

Oaksford, Mike and Chater, Nick. 2007. *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.

—. 2020. “New paradigms in the psychology of reasoning.” *Annual Review of Psychology* 71:305–30.

Proust, Joëlle. 2013. *The philosophy of metacognition*. Oxford University Press.

Rips, Lance. 1994. *The psychology of proof: Deductive reasoning in human thinking*. MIT Press.

Simon, Herbert. 1976. “From substantive to procedural rationality.” In T.J. Kastelein, S.K. Kuipers, W.A. Nijenhuis, and R.G. Wagenaar (eds.), *25 years of economic theory*, 65–86. Springer.

Slovic, Paul, Finucane, Melissa, Peters, Ellen, and MacGregor, Donald. 2007. “The affect heuristic.” *European Journal of Operational Research* 177:1333–52.

Thaler, Richard. 1990. “Anomalies: Saving, fungibility, and mental accounts.” *Journal of Economic Perspectives* 4:193–205.

Titelbaum, Michael. 2012. *Quitting certainties: A Bayesian modeling framework modeling degrees of belief*. Oxford University Press.

Todd, Peter and Gigerenzer, Gerd. 2012. *Ecological rationality: Intelligence in the world*. Oxford University Press.

Zhu, Jian-Qiao, Sandborn, Adam, and Chater, Nick. 2020. “The Bayesian sampler: Generic Bayesian inference causes incoherence in human probability judgments.” *Psychological Review* 127:719–48.