Lancieri F., E. Posner and L. Zingales 2022. The Political Economy of the Decline of Antitrust Enforcement in the United States. No. w30326. National Bureau of Economic Research. https://www.nber.org/papers/w30326.

Peruzzi E. 2022. Economic models in the courtroom: a methodological approach (May 2022). Available at https://ssrn.com/abstract=4213542.

Shapiro C. 2021. Antitrust: what went wrong and how to fix it. *Antitrust* **35**, 33–45.

Wollman T. 2019. Stealth consolidation: evidence from an amendment to the Hart-Scott-Rodino Act. *AER: Insights* **1**, 77–94.

Edoardo Peruzzi is a PhD candidate in Economics in the programme of the Tuscan Universities. His research interests lie at the intersection between philosophy of economics, antitrust economics and history of economics. His current project concerns how economics is employed in antitrust court cases. URL: https://sites.google.com/view/edoardoperuzzi/home-page.

Joaquín Paseyro Mayol is a PhD candidate in Economics in the programme of the Tuscan Universities. His research interests are in labour and public economics, with applications in health, education and development economics. He is currently studying the impact of highly effective contraceptives on women's educational and employment outcomes. URL: https://sites.google.com/view/joaquinpaseyromayol/home-page.

Theory and Credibility, Scott Ashworth, Christopher Berry and Ethan Buena de Mesquito. Princeton University Press, 2021, 280 pages.
doi:10.1017/S0266267123000056

Here's the top line: This is the best book on the philosophy of social science I've read in a long time. If you're interested in the philosophy of social science, social-scientific methodology, or issues of modelling and explanation, you should immediately make your way to the nearest purveyor of fine books and buy yourself a copy or three.

The book's philosophical core is its discussion of idealization and robustness. To be sure, these are familiar issues for philosophers of economics. But *Theory and Credibility* is like a good steak – its great virtue is the quality of its execution. In particular, its authors ably link issues of philosophical substance with concrete examples that are presented simply enough to be accessible to philosophers with little formal economics or political science training, but in enough detail to motivate those issues to practicing social scientists and link them tightly with actual social scientific practice.

By their own lights, Ashworth, Berry, and Buena de Mesquito (hereafter ABB) are interested in explaining to theorists and empirical social scientists what, exactly, the other party is up to, and why they do the things they do. I am neither a theorist nor an empirical social scientist, but I came away with a great deal of insight into both sides of the social science enterprise.

The book is divided into two parts, bookended by a brief Introduction and a Conclusion. The first part, 'Foundations', comprises the first five chapters and provides a comprehensive conceptual framework for thinking about how theory and empirical methods help us discover things about the world. The second,

'Interactions', applies the framework to show how particular interactions between methods and theory can help extend our ability to answer social-scientific questions.

Chapter 2 is the first really substantive chapter of the book; it introduces ABB's framework, which they then elaborate over the next three chapters. The framework comprises three parts – a (theoretical) model, an empirical research design, and a target phenomenon – and three relations among them: *commensurability* of the model's implications and the estimates of particular quantities important to the research design (Chapter 3); *similarity* (in the relevant sense) between the model and the target (Chapter 4); and *similarity* between the research design and the target (Chapter 5). The similarity and commensurability relations provide a framework for understanding the specific theoretical and empirical techniques social scientists use to understand interesting phenomena.

Commensurability is just the idea that the implications of the model and the estimates spat out by the research design are plausibly *about the same thing*. ABB's elaboration of commensurability in Chapter 3 concerns the idea that at its core, social science research is about *ceteris paribus*, or "all else equal" relationships. That is, whether or not I choose to delegate choices to a better-informed agent depends, all else equal, on how my preferences differ from theirs. It might also depend on other things – the agent might be a moral reprobate, and thereby untrustworthy. We care about *ceteris paribus* relationships in theory as a consequence of the fact that formal models involve fixing some primitives in order to draw implications about mechanisms. Particular models don't tell us anything about what changes when their primitive assumptions are specified differently. The point, then, is that making sure that commensurability holds – that the implications of our theory are the same thing our empirical research design is talking about – requires that both our theory and empirical research design hold some things fixed, which requires that our social-scientific inquiry deals primarily with all-else-equal relationships.

Perhaps the most philosophically rich part of the book is the discussion of the similarity relation in formal theory in Chapter 4. Suppose you're interested in whether (say) economic factors cause people to become terrorists, or whether the 'perception gap' – that is, women systematically underestimating their quality as electoral candidates – is responsible for women's underrepresentation in elected offices, or why Congress votes more along party lines now than in the past. Surely it's not enough to show *that* women are underrepresented in elected office, or *that* people become terrorists, or *that* Congress votes along party-lines more often now than in the past (which is not to denigrate the importance of descriptive statistics!). We want to know *why*. And to answer the *why*-question, we'll want to identify a *mechanism*, the feature of the world responsible for the outcome we care about (49). The problem is that we can't directly observe causal relationships, except in some *highly* artificial settings of the sort exploited by experimental economics (Guala 2005). Instead, we have to *infer* causal relationships. There are three things we might want to know in order to do that. One thing we want to know is what we might expect to see if, in fact, that mechanism operates within the real world. This is the job of a theoretical model, and articulating it is the job of Chapter 4. (Another thing we want to know is whether the mechanism we've identified is the thing that actually explains why

our target system behaves the way it does. This is the job of an empirical research design, which ABB talk about in Chapter 5.)

How does a model tell us about a mechanism that might explain the thing we're interested in? It has to be relevantly *similar* to the target system. What does that mean, and how do we tell? After all, we can't directly observe causal relationships in our target phenomena. If we could, none of this would be a problem. Social science would be a lot easier. The central problem is that models *idealize*. They say things that are, strictly speaking, *false* about a target system. For example, it is well-known that the behavioural assumptions of rational choice theory are subject to numerous objections, which, if true, tell us that RCT doesn't do a great job describing the behaviour of actual people (Kahneman and Tversky 1979; Kahneman 2011). How, then, does saying false things about a target system help us get a better handle on its behaviour? What's more, how can we tell whether a model is similar to the target system, without having information about the target system that would obviate the need for the model in the first place? Obviously, these are not new questions in the philosophy of social science (e.g. Weisberg 2013; Rice 2021), but they raise important epistemic and practical questions. Pamuk (2021) has recently articulated some further potential consequences: changes to a model so its implications fit better with empirical observations do not guarantee that model better represents the underlying phenomena.

In ABB's language, a model has to be *relevantly similar* to the target system, which is to say that exploring the features of the model can tell us about important features of that target system (47). If Pamuk's analysis is right, similarity is essentially a black box – we cannot know anything about how similar the model is to the target. The relationship between the model's outputs or implications and our observations of the target system are no guarantee that similarity holds. Her move is to argue that because we cannot know how accurate the model is, its specification is a matter of the values of the scientific and policy community (Pamuk 2021: 38–40). One great virtue of ABB's analysis is that it shows that our epistemic situation isn't *nearly* so bleak, and that the similarity relation can help us understand what makes some models more accurate than others. Another is that it does so by appeal to details of scientific practice in economics and political science, which serve as counterexamples to Pamuk's sceptical story.

So, the basic move, in a theoretical model, is to specify a set of relationships between quantities of interest, and then argue that the mechanisms represented by those relationships in the model are likely to show up in the real world, by articulating the ways we ought to expect the real world to be if it works the way the model says it does. One example of this from *Theory and Credibility* is a set of models of *delegation* (49–53). When do individuals delegate decisions to better-informed others? There are two stylized facts of interest: first, uninformed principals are more likely to delegate to agents when the agents' preferences are close to the principal's, or when there's a lot of uncertainty about the relationship between policies and outcomes (50).

One classical model specifies a particular utility function called a 'quadratic-loss' function, according to which utility decreases according to the square of the distance

between preferred policy and the chosen policy. This is meant to represent risk-aversion, in order to formalize the intuition that risk-averse principals will be more likely to delegate, because the outcomes of an agent's policy preferences are more of a 'sure thing' (50).

However, it turns out that the result generalizes beyond the simple quadratic-loss utility function, which shows that risk-aversion is an auxiliary assumption, and not part of the core mechanism. Essentially, this tells us that willingness to delegate depends on how far from the principal's optimal policy the agent's favoured policy option might be, rather than on the principal's risk aversion, or the specific shape of the principal's utility function. It also relieves the theorist from having to argue that actual agents act as the quadratic-loss utility function says they should. This means that a larger set of ways the world can be are compatible with the model's outputs.

What this process of reasoning using the model shows is that even leaving aside empirical research design, theorists can make significant progress in figuring out what variables are important and which mechanisms are actually represented, and which auxiliary assumptions need not be made. What makes these changes to models of delegation *improvements* can be understood in terms of the similarity relation. The model becomes a better representation of the phenomenon of interest because we can see that something we might have thought to be an important constituent of the mechanism we're interested in is, in fact, an auxiliary assumption that we can (and should) jettison. Showing that a particular relationship generalizes amounts to showing that it doesn't actually matter whether a model is accurate along some dimension (in this case, whether the quadratic-loss utility function actually describes agents' preferences). So far from the similarity relation being an explanatorily inert black box, it in fact guides the process of theoretical inquiry.

This example helps highlight what I think is philosophically distinctive about ABB's approach, which is to cast the similarity relation in explicitly pragmatic terms, and thereby link it to the idea of robustness analysis. Traditionally, 'robustness' has a few different meanings in social science. The most notable – or at least, the one economists on Twitter are most likely to complain about – is specifying research designs in various ways in order to make sure that a result reflects the underlying reality rather than some 'chance or chicanery' (69). But another way to think about robustness analysis is as identifying the ways of specifying a mechanism relative to which a particular implication is invariant. By showing that the relationship between risk aversion and a principal's willingness to delegate varies across changes in the functional form of the principal's utility function, we can infer that the more general model is more similar to the target phenomenon. Similarly, I think the example helps show that robustness and de-idealization aren't exactly the same thing (Lisciandra 2017). The greater similarity to the target system of the more general model of delegation isn't due to the fact that it represents agents' utility functions any more accurately than the quadratic-loss function. It's not as though it uses a general theory of agents' utility inferred from surveying actual agents. Rather, its greater similarity to the target system is due to the fact that it omits auxiliary assumptions. I don't think that this is a revolutionary new view of robustness or similarity. But ABB's

pragmatic account of similarity, combined with the clarity of the examples, help make clear the specific epistemic contributions robustness analysis makes to a theoretical model.

Of course, even a plausible model can be mistaken about what mechanisms actually drive the phenomena we care about. The world is a complicated place, and *lots* of different mechanisms might play a part in the results we see. And we might want to know the extent to which a particular mechanism is responsible for the phenomenon of interest. Chapter 5 lays out what ABB call the 'Elements of a Research Design' (ERD), which articulates in detail the similarity relation between a target system and a research design. This is the longest chapter of the book, and is complex enough that I won't be able to do it justice. Still, there are a few philosophically interesting features I'd like to talk about. The ERD has four parts: an empirical strategy, an argument about measurement validity, an argument about substantive identification, and a confidence-building strategy. The empirical strategy basically encompasses an estimand – an empirical relationship that the research design outputs – a statistical procedure, chosen for its fit with the underlying structure of the phenomenon in question, and data, which should be amenable to uncovering the statistical relationships of interest. Measurement validity concerns whether the data set and estimand are in fact about what we think they're about. For example, measures of political polarization that focus on policy preferences may not adequately capture the sense of 'polarization' relevant to studying contemporary politics (Fiorina *et al.* 2011; Iyengar *et al.* 2019). Substantive identification is the attempt to show that the assumptions underlying some particular statistical technique actually hold. And the last bit, 'confidence building', is about making sure that 'findings are unlikely to be due to chance or chicanery' (72). While I lack the space to talk about it in detail, ABB's discussion of statistical methods is incredibly accessible – I have yet to see a clearer discussion of the underlying logic of difference-in-differences and regression discontinuity research designs. Rather, I'll concentrate on their discussion of the similarity relation. ABB argue that each of the four components of the ERD helps us articulate the similarity relation between the target phenomena and the research design. Most important are questions of measurement validity and substantive identification, which give substance to the more general similarity criterion for research designs.

Part II of the book elaborates on and extends ABB's framework. It is in some ways the most impressive part of the book. It comprises five chapters: 'Reinterpreting', 'Elaborating', 'Distinguishing', 'Disentangling' and 'Modeling the Research Design'. The first four of those five are about the ways theory and empirical research designs can interact in order to enrich our understanding of phenomena we might be interested in. Each of these chapters helps us see how the similarity relation is useful for understanding actual social science research. I'm going to concentrate on two of the chapters – '"Reinterpreting' and 'Distinguishing' – as they help show how social scientific research with idealized models can answer traditional epistemic challenges.

Chapter 6, 'Reinterpreting', argues that one role of theory is to, well, reinterpret empirical findings. ABB use the literature on party effects in Congressional voting to show that multiple mechanisms are compatible with the empirical result that

members of Congress vote along party lines more often than they used to. One mechanism is that the party exercises more control over individual members. Here, the party is conceived of as an agent, exerting pressure on members' voting behaviour, generally in exchange for support in re-election campaigns. Another mechanism is sorting. It may simply be that members of Congress who are antecedently, ideologically disposed to vote for some bill or other will be more likely to run as members of a particular party (140). In these cases, theoretical models can help articulate different mechanisms that might be responsible for a phenomenon whose existence is agreed upon.

Chapter 8, 'Distinguishing' also does exactly what it says on the tin. The point here is that, as in the case above, multiple mechanisms can be responsible for the same effect. In order to figure out which of a number of plausible mechanisms is responsible, we can try to find *distinguishing* implications – that is, things implied by one model but not by another, which we can then test for. If party-line voting is due to party control, then ideologically similar members of different parties should vote differently on particular bills. On the other hand, the sorting mechanism entails that ideologically dissimilar co-partisans – think, in the American context, of Senators Sanders and Manchin, or Hawley and Romney – will vote differently on some bills (189). As it happens, ABB appeal to a study by Ansolabehere *et al.* (2001) that shows ideologically similar members of different parties tending to vote differently on bills, which is, all else equal, an indication that the party-control mechanism helps determine voting behaviour – or, put differently, that party-line voting isn't due entirely to ideological sorting (190–191).

These two chapters emphasize that the relationship between theory and confirmation is more complicated than philosophers often appreciate. I think the most philosophically important feature of ABB's picture is that the similarity relation is *pragmatic*. The fundamental modelling question is about what we can learn about the real world from the model. The possibility of reinterpreting an empirical result by drawing out the implications of different, competing models to see which better explains the empirical result shows us that the compatibility of an empirical result with the implications of a formal model need not serve as confirmation of the story the model tells about what is going on in the world. Similarly, elaborating on a model by teasing out further implications of a mechanism can help distinguish it from other mechanisms. One nice way of thinking about this section is as a taxonomy of robustness analyses, each of which helps improve similarity. That is exactly what Pamuk claims cannot be done. The basic mechanics of reasoning with models can help us figure out whether a model is a good representation of a social system – that is, whether it is relevantly similar to that system – by figuring out which features of the model are the important ones, and how we might expect a particular mechanism to play out in reality.

Anyway, it's a fabulous book and I can't say enough good things about it.

Kirun Sankaran 🔾
UNC-Chapel Hill
Email: kirun@email.unc.edu

References

Ansolabehere S., J. Snyder, Jr. and C. Stewart III 2001. The effects of party and preferences on congressional roll-call voting. *Legislative Studies Quarterly* **26**, 533–572.

Fiorina M.P., S.J. Abrams and J.C. Pope 2011. *Culture War? The Myth of a Polarized America*. New York, NY: Pearson.

Guala F. 2005. *The Methodology of Experimental Economics*. Cambridge: Cambridge University Press.

Iyengar S., Y. Lelkes, M. Levendusky, N. Malhotra and S.J. Westwood 2019. The origins and consequences of affective polarization in the United States. *Annual Review of Political Science* **22**, 129–146.

Kahneman D. 2011. *Thinking, Fast and Slow*. New York, NY: Farrar, Strauss, and Giroux.

Kahneman D. and A. Tversky 1979. Prospect theory: an analysis of decision under risk. *Econometrica* **47**, 263–292.

Lisciandra C. 2017. Robustness analysis and tractability in modeling. *European Journal for the Philosophy of Science* **7**, 79–95.

Pamuk Z. 2021. *Politics and Expertise*. Princeton, NJ: Princeton University Press.

Rice C. 2021. *Leveraging Distortion*. Cambridge, MA: MIT Press.

Weisberg M. 2013. *Simulation and Similarity*. Oxford: Oxford University Press.

Kirun Sankaran is Teaching Assistant Professor of Philosophy at UNC-Chapel Hill. He works on issues at the intersection of political philosophy and the philosophy of social science. His 2021 paper 'Structural Injustice and the Tyranny of Scales' argues that insights from multiscale modelling in the physical sciences can help us articulate what is distinctive about 'structural injustice'. He is currently working on projects on Bernard Williams, Thorstein Veblen, and conceptual issues regarding causal inference in social network models.