Formal philosophy as modelling

Abstract

It is increasingly common to claim that modelling is a useful method for philosophy. Williamson, Hartmann, Leitgeb, Godfrey-Smith, and Paul have all made versions of this claim; the first three for formal epistemology (FE), the latter two for metaphysics. This paper has two aims. The first is to explain what talk of "modelling" in philosophy could mean, by applying the philosophy of scientific modelling to work in formal epistemology. The second is to illuminate the methodology of formal philosophical work, to an audience of general philosophers and those doing formal work themselves. Beginning with a characterisation of modelling and its methodological constraints, I establish the initial plausibility of saying that FE involves modelling. I then address the core difference between scientific and philosophical modelling: normativity. I show that we can extend the scientific account of modelling to cover normative work. I then turn to methodological implications for formal philosophy, arguing that certain inferences are ruled out once we acknowledge that we are modelling. I close with a reflection on how modelling can help or hinder our normative inferences.

1 Introduction

Talk about "modelling" as a method of philosophical inquiry is increasingly prevalent, across various philosophy subfields. Williamson (2006, 2017) has named modelling as important method for a certain style of philosophy; what we might call scientific or mathematical philosophy. He defends modelling as a tool for developing clear arguments (2006, pp. 186–7), and as a major source of philosophical progress (2017, p. 8). Similarly, Stephan Hartmann has done much work using and discussing models in philosophy (Bovens and Hartmann, 2003; Eva and Hartmann, 2019), and Hannes Leitgeb (2013, p. 273) agrees that modelling is a method for building inductive strength in an argument. Peter Godfrey-Smith (2006, 2012) and L.A. Paul (2012) have discussed modelling as a practice in metaphysics. Michael Titelbaum (2012) describes the project of his book as providing a framework for building models in formal epistemology.

In all these cases, the talk of modelling and model-building is an analogy with the common-place scientific practice of indirect inquiry using idealised representations. We can situate this talk as part of a wider naturalistic turn in thinking about philosophical methodology. The first aim of this paper is to

provide a clear explanation of how the scientific methodology of modelling can work in philosophy, by focussing on the specific case of epistemology.

This will be valuable to philosophers engaging with formal philosophical work "from the outside," for example mainstream epistemologists engaging with the increasingly popular formal epistemology literature. But it is also valuable to formal philosophers themselves. There are differences between the philosophical and scientific cases, which are in need of explanation if the talk of "modelling," and the claim to methodological legitimacy it underpins, are to succeed. I will focus on one crucial difference: philosophy is often normative, and the objects being called "models" in philosophy serve normative purposes; while science is not typically normative, and models in science serve descriptive, explanatory, or predictive ends.[1] How does this difference influence the claim that we are (sometimes) modelling in philosophy? I will argue that normative work *can* be considered modelling, though there are some unique considerations in the normative case concerning the role of idealisation.[2]

Though I will not comment on ethics directly, an account of normative modelling in philosophy will, I think, be of great use to moral and political philosophers. There is a long history of discussion about the role of idealisation and abstraction in ethical theory, (e.g., O'Neill, 1987), and how it relates to the distinction between ideal and non-ideal theory (e.g., Mills, 2005). As some have noted (e.g., Hancox-Li, 2017) there are obvious parallels between idealised and abstracted ethical theorising and scientific modelling. My account of normative modelling and its role in another normative field (epistemology) will hopefully be of use to that discussion, though additional work will be required to reap the benefits I promise here, as my focus here will be on *formal* normative modelling and much of the relevant ethical work is not formal. Of course it will also be of direct use in explaining what formal ethicists are up to.

Second, the growth of formal epistemology (FE) has led to a spread of its ideas into more mainstream philosophy without a corresponding dissemination of its methods. Many philosophers now consider the attitude of partial belief (or degree of belief) to be an important topic in epistemology; typically in the form of "credence", a particular mathematical representation of that attitude as a number between 0 and 1. I offer an explanation, for a wide philosophical audience, of what formal philosophers are up to and thus how one should regard objects like credence. My conclusions are cautioning: normative work using models is complex, and a number of inference-patterns familiar from other parts of philosophy do not work well here, including certain realist inferences, and reasoning by counterexample.

In section 2, I will describe my target more completely by outlining one ex-

[1] Note that our subject here is not representational models of communities obeying norms, or of how norms might emerge, such as those studied in the literature on the social evolution of morality. I am here interested in models whose purpose is the generation of normative claims.

[2] Colyvan (2013) has discussed the role of normative assumptions in formal models such as Bayesian decision theory and logic. My project differs in that I provide an account of the content of normative models, and their relation to the world, along with more explicit methodological conclusions for formal epistemology in particular.

ample of what formal epistemologists call a model. In section 3, I present a loose characterisation of scientific modelling, and some lessons from the philosophy of science literature on it. These are deployed in section 4 to characterise formal epistemology as modelling in a first pass, and then extended in section 5, where I present my account of normative modelling. In section 6 I consider alternative explanations of what formal epistemologists could be doing—modelling is just one methodology among many, and it is helpful to distinguish between them both to delineate the alternatives and highlight the implicit contrast between my account and the commonplace view of philosophical methodology. I then turn in section 7 to some methodological considerations for FE, given that we are modelling. A key part of the discussion is a consideration of when normative conclusions drawn from models are "secure"—well justified by the model—given the dependence of modelling on idealisation. Section 8 concludes.

2 The target

Though my conclusions apply broadly to modelling in formal philosophy, I will discuss formal epistemology (FE) and specifically the attitude of graded or partial belief, for clarity.

Let us start with an example of the kind of structure that I want to call a model. Consider some common modes of inquiry in formal epistemology, concerning partial belief.

- *The nature of rational partial belief.* Here, we translate norms of rationality into a formal (i.e., mathematical) setting, and use the precision this affords to draw conclusions about the implications of those norms for the structure of our attitude of partial belief.
- *The norms of rationality.* In another mode, the norms themselves at issue. We examine the plausibility of putative norms by translating them into a formal setting, deriving some results, re-translating the results back into ordinary language, and testing them against firmly held intuitions.
- *Decision-making.* Pairing norms of rational belief with norms of rational desire allows us to derive rules for selecting one option from a menu of possible acts. Again, we might explore the implications of norms we already hold or test the implications of putative norms.

In each case, we start with an initial question/problem framed in natural language. Some principles of rationality governing the agents involved are chosen for investigation. These are translated into a formal language capable of representing agents, propositions, beliefs, and so on (as necessary). Constructing this formal apparatus typically requires introducing additional structure, that is not motivated by the initial question but is internal to the process of representing it mathematically. The formal setup is then studied, and conclusions are drawn. Finally, these formal results are translated into conclusions about partial belief or decision-making.

Here is a specific example. We might begin with some observations about our subjects: people have partial beliefs. (Partial beliefs are often communicated in the language of likelihood, and are often referred to as "comparative likelihood judgements", or "comparative confidences".) We observe people making statements about their confidence in various judgements or making comparative judgements about two propositions they avow to believe, such as "I am fairly certain that Brexit will turn out to be a disaster" or "I am more confident that it will rain tomorrow than I am that Boris Johnson will make a good Prime Minister."

Under good conditions, we observe that these partial beliefs have the following properties. Partial beliefs are *monotonic*: we believe logically weaker propositions to a greater degree than stronger. I believe "it will rain on Monday or Tuesday" more than I believe "it will rain on Monday". They are *separating*: we can "factor out" common prospects when making comparisons. If I regard it as more likely it will rain on Monday than on Tuesday, then I also regard it as more likely that it will rain on Monday or Wednesday than on Tuesday or Wednesday. Finally, comparative partial beliefs are *transitive*: if I believe it is more likely to rain on Monday than Tuesday, and more likely on Tuesday than Wednesday, then I must believe it is more likely to rain on Monday than Wednesday.

In the first two cases, "under good conditions" means something like "when we're aware of, and think consciously about, the logical relations between the relevant prospects." In the third case, it means something like "when the initial two pairs of comparisons are considered together." These patterns strike us as reflecting something about the logic of partial belief, and as we begin to theorise the attitude we conclude that believing in a way that doesn't fit have properties would be doing something wrong. These therefore become putative norms for rational partial belief. This is our topic of study.

What distinguishes formal epistemology is the decision to represent this attitude with some mathematical object. In this case, it is common to represent partial belief with a binary relation, denoted $\succeq$, which encodes an agent's comparative judgements.[3] Writing R for "it will rain tomorrow" and B for "Boris Johnson will make a good Prime Minister", $R \succeq B$ represents the fact that the agent is more confident that it will rain tomorrow than they are that Boris Johnson will make a good Prime Minister. I will refer to the mathematical relation $\succeq$ as the "credibility" relation. Credibility is defined on a Boolean algebra, which is a set of propositions Ω that is closed under $\models$, an implication relation. This implication relation can be defined as $X \models Y \iff X \vee Y \models Y \iff X \wedge Y \models X$, for classical con/disjunction (Bradley, 2017).

What makes credibility stand for partial belief is that we then endow it with mathematical properties that correspond to the attitudinal properties we settled on above. We stipulate that it is monotonic over $\models$, $\vee$-separable (Joyce (1999, p. 91) calls this "quasi-additive"), and transitive. We take these mathematical

[3] A note for readers who have encountered decision theory or microeconomics, but not formal epistemology: $\succeq$ does not represent *preference*, it represents comparative partial belief. Preferences are not discussed anywhere in this essay.

properties to represent the norms that we theorised for the attitude of partial belief, just as credibility represents that attitude.

Binary relations are not particularly easy to work with, and so formal epistemologists typically make progress by using a "representation theorem". A representation theorem is a mathematical argument showing that a binary relation $\succeq$ can be represented by a real-valued function $F : \Omega \to \Re$, where this means just that $F(X) \geq F(Y) \iff X \succeq Y$. Such a theorem typically specifies the form of the function, and some uniqueness conditions for it. Under certain conditions, credibility can be represented by a probability measure. I will use common jargon and call this a "credence" function, denoted P. For credibility to be represented by a credence function, it needs to have certain mathematical properties. The details vary depending on the particular representation theorem, but if we examine important theorems which result in unique probability functions—Villegas's theorem and Joyce's theorem—we observe two kinds of conditions. The first are precisely those normative conditions discussed above: monotonicity, transitivity and separability. The second includes, for example, the requirement that credibility must be *complete*: for any two propositions X, $Y \in \Omega$, they must be related in some way by $\succeq$: either $X \succeq Y$, $Y \succeq X$, or both.[4]

We will return to the details of these representation theorems below, but for now we note that completeness is not a particularly compelling norm for partial belief. It requires that agents can compare any proposition, no matter how obscure, with every other proposition in terms of its comparative likelihood. It is therefore common to regard this second group of requirements on credibility as *non-normative*. The model therefore comes to include both normative and non-normative assumptions about its mathematical consituents.

This mathematical apparatus is intended for normative use, however. Philosophers who work with such tools are interested in the rational structure of the attitude of partial belief; they are interest in (amongst other things) whether or not we ought to have probabilistic partial beliefs. So, using the representation of partial beliefs by probability functions (under certain conditions), they seek to derive a norm: if your partial beliefs cannot be so represented, then you are irrational. (This norm is called Probabilism.[5])

Two comments are important at this early stage about normativity. First, the primary mode of normativity operant here is that of evaluation. This is a standard against which we are measured; it is out of reach but linked in important ways to our actual capacities. There is a linked, secondary mode of normativity—prescription, or action-guidance—that is common in some parts of decision theory. I will focus on the evaluative mode here.

Second, there are two common ways that such formal normative work takes

[4]More fully: these two theorems require credibility to be monotonic, separable, transitive, complete and continuous. The Boolean algebra must also be complete and atomless. Bradley (2017) discusses each assumption.

[5]There are, of course, other (better) ways to argue for this norm. I don't claim this is how it ought to be done, but grant me that it is sometimes defended in this way for the purpose of the discussion.

place in epistemology. We might work constructively, introducing and defending each assumption about credibility in turn and concluding with the norm of Probabilism. The defences are typically that the assumptions are themselves norms (e.g., monotonicity as I presented it above), or that they are true descriptions of the attitude, or that they are harmless structural requirements—mathematical conditions that don't represent anything but are useful to get the discussion moving.[6]

Alternatively we might work critically, by starting with a formal apparatus and criticising it for making the wrong ruling: either it declares something to be bad that is in fact good, or vice versa. (I'm using "good" and "bad" here for the two valences of the relevant norm, e.g., rational and irrational.) In decision theory, these critical engagements often involve particular choice situations, like the Allais (1953) and Ellsberg (1961) choices. A particular decision theory rules the Allais choices irrational but, says the critic, such choice are intuitively *rational*. Thus, that decision theory is flawed and requires alteration.

Some of the above is obviously similar to scientific modelling—the use of mathematics in a representational role, the presence of idealisations. But some is peculiar to philosophy. So, are these models? If so, how do they work and does their methodology come with the same constraints and benefits as descriptive modelling?

3 Scientific models

Let us begin with a review of scientific modelling, and the methodological lessons we have learned from five or six decades of philosophical study of modelling.

'Model" is one of those unhelpful terms that is used to mean many different things, so I want to begin with a common meaning that I *do not* intend to use: the meaning logicians give to the term. Roughly put, logicians use "model" to mean an interpretation that satisfies a set of sentences. An interpretation is here an assignment of semantic values to the basic vocabulary in use. This semantic sense of "model" takes it to pick out certain *mathematical structures*. Some philosophers of science (e.g., Suppes (1969)) have argued that this meaning of "model" is the same as, or should be used to explicate, the workaday use of "model" in scientific practice. This is a view which is associated with the once popular "semantic view" of a different scientific item of interest, the theory. I will not be using "model" in this sense, and in that I will diverge at the outset from some (like Paul (2012)) who have discussed modelling in philosophy.

The way I use the term "model" is broadly consistent with a philosophy of science tradition that includes Giere (1988, 2004) and Cartwright (1989), as well as the many others cited below, and in terms of the nascent literature on philosophical models it is more or less how Godfrey-Smith (2006) uses the

[6]Not everyone is so cavalier, of course! Joyce thinks "that the Achilles heel of Savage's theory is its dependence on structure axioms that cannot be satisfactorily explained away" (Joyce, 1999, p. 98) —a conclusion quite close to mine, although not presented in anything like the same way.

term.[7] If you typically think of models as set-theoretic structures you will need to take this section as stipulating a new meaning for that term.

So what is a model? Here are three examples, to ground your intuitions as I introduce the theoretical account. Some models are material objects, like the molecular structure models used by chemistry students. Modelling kits, such as the MolyMod system, come with coloured balls representing elements (red for Hydrogen, black for Oxygen), and grey connecting rods representing chemical bonds (short and stiff for single-valence, long and bendy for double-valence). With these kits, students build models of simple molecules like H_2O, and more complex polymers like PVC. We call the real-world system under study the target, and the plastic object the model. The model of H_2O involves one red ball connected by two short grey rods to two smaller black balls, in a wide V shape. The student learns about the structure of the molecule, H_2O, by examining the plastic model.

More commonly, models are theoretical rather than material. The Bohr Model of the atom is a classic example: Bohr imagined the Hydrogen atom as an orbital system consisting of a central positively charged sphere orbited by a distant, negatively-charged sphere. The centre represents the nucleus, the orbiting sphere represents the electron. The "electron" is in a circular orbit, and only certain orbits (with specific orbital distances) are allowed. What is the model in this case? It is described by a series of written statements (like those above, together with Bohr's "rules" for electron orbits), often accompanied by equations (e.g., $L = n\hbar$, specifying the angular momentum of the orbiting electron) and perhaps illustrated with a diagram. But the system we are investigating when we use the Bohr Model is not identical with any, or all, of these physically instantiated parts; it is what those descriptive elements specify (Mäki, 2009, p. 33). There are several philosophical accounts of what such "theoretical models" are, but for now we need only note that, whatever they are, they are not material objects.

Finally, some models have no target. Architectural plans for buildings which will never be built are models, as are theoretical models for ether or phlogiston, substances which do not exist. In modern quantum field theory, "ϕ^4 theory" is a simple, intuitive model which has been extensively studied despite being known not to correspond to any physical system (Frigg and Hartmann, 2018). So, a philosophical account of models must rely neither on a concrete model system, nor on a concrete target system.

Philosophers of science have developed a rich literature on the representational function of models, their ontology, epistemology, and implications for scientific realism (see Frigg and Hartmann, 2018). I will here draw attention to a few lessons learned in this literature, for comparison with the practice of

[7] I will not attempt a classification of all of those writers about philosophical modelling mentioned in the introduction. For one thing, many writers switch between different senses of the term model in the same paper. This is the case with Paul (2012), who appeals to the work of Godfrey-Smith (2007) alongside semantic-view authors whose views Godfrey-Smith explicitly repudiates.

formal epistemology. [8]

(1) Modelling is characterised by indirect inquiry (Giere, 2004; Godfrey-Smith, 2007; Weisberg, 2007b). Instead of studying the natural system, modellers describe and investigate a "model system" which is the primary target of their investigation. The model system is taken to (partially) represent the target natural system. Modellers then infer facts or generate hypotheses about the target system based on their investigation of the model system. (In cases where the model is target-less, they are still thought to be representational in a sense to be discussed later.)

(2) Models present an idealised and distorted picture of the target system (Frigg and Hartmann, 2018; Weisberg, 2007a). Many real-world systems cannot be investigated directly, due to incomplete theories or severe computational complexity. To make progress, scientists simplify the system under investigation, by changing or leaving out aspects of the real system. They work to identify the features of the system most salient to their investigation (Weisberg, 2013, p. 4). The frictionless plane is a classic example: no real surface is frictionless, but it is fruitful to take a surface to be frictionless when investigating inertial motion of objects on an inclined plane.

There is an extensive literature in idealisation in science; I will note two distinctions drawn in that literature for use here. There are different kinds of idealisations: Galilean and Aristotelian (Frigg and Hartmann, 2018). Galilean idealisations introduce deliberate distortions to some properties of the system under investigation. For example, the friction of the plane is deliberately changed in the representation. Aristotelian idealisations leave out features of the system that are not relevant to the problem being studied, to allow us to focus on or isolate a limited set of properties. For example, a population growth model considers only the rate of reproduction and predation of organisms and leaves out all their other properties.[9]

There are also different motivations for idealisations (Musgrave, 1981). A modeller might take a property to be negligible, believing that for the purposes of the current investigation it will make no difference to distort/exclude it. For example, we might consider falling objects and idealise by assuming there is no air resistance because we believe it to be of negligible importance. Another way of putting this is that the idealisation functions well when it is true that the effect of air resistance is small, so that the model's claim that air resistance is zero is approximately true.[10] Alternatively, the modeller might know that the property is not negligible in all cases but want to model only those cases where it is so. Musgrave calls this a domain idealisation: it justifies itself "automatically" by restricting the class of cases the model applies to. Finally,

[8] While they are not without opposition, I aim to use only "mainstream" views in the philosophy of modelling. I also do not attempt to provide anything like a complete bibliography on each point here. Rather I cite a few recent sources, with good references in each for the interested reader.

[9] Some authors call Aristotelian idealisations "abstractions," though usage is by no means uniform.

[10] I don't want to be committed to an approximate truth account of idealisation here; I am merely presenting some ways idealisations are thought of.

the modeller might think that there are no cases where the property is negligible but distort/exclude it anyway because its presence in the model makes things too complex to handle. Musgrave calls this a heuristic idealisation, and presents it as part of a process of inquiry: we simplify the model by setting air resistance to zero now, with the hope that once we have established the model we can factor air resistance in later. Note that negligibility, domain-restriction and heuristic necessity are species of justification—the same idealisation can be justified in each way, depending on the modeller and the circumstances.

(3) Models are built for a purpose, and so perform well only within a restricted domain of applicability (Parker, 2009; Teller, 2001; Weisberg, 2007b). "Purpose" consists of what you're modelling (e.g., ants rather than bears) and what you're trying to do (e.g., study group coordination). This establishes the basic domain of the model (it is a model of ant coordination). As Wimsatt (2007, p. 15) points out, models are often used to isolate particular mechanisms or concepts for study. This purpose motivates the idealising assumptions, which may further restrict the domain of applicability as discussed above. I'll refer to the combination of purpose and domain as the model's scope.

The purpose-driven nature of modelling means that model-based sciences often contain multiple, disagreeing models of the same phenomena. Teller illustrates this with an example of two models of water. The first is interested in the flow of water and wave propagation, and it models the liquid as a continuous incompressible medium. The second is interested in explaining diffusion, say of a drop of ink in water. It models water as a collection of discrete particles in thermal motion. Each is similar to water in the respects that are relevant to its purpose, but the two models look very different (Teller, 2001, p. 401). Neither should be thought to provide a definite characterisation of water, and our understanding of water is enhanced by having both available.

4 The methodology of modelling

The foregoing characteristics of modelling and models lead to certain methodological constraints for this kind of science. Idealisation is the lifeblood of modelling, but while it helps scientists make progress in investigations of complex systems, it introduces limitations. As Levins (1966) put it, modelling involves an inherent three-way trade-off between precision, realism and generality of scope.

On the realism front: models contain artefacts, properties of the model system that are not representative of any real feature of the target system but instead emerge from the representational choices of the modeller or the idealisations in the model. Good modellers must identify artefacts and ensure that they aren't imputed to the target. If there is an underlying fundamental theory (as is often the case in physics), this can help to identify artefacts. Another method for identifying such effects is sensitivity analysis.[11] This is a method

[11] Also called stability analysis, it is closely related to what Weisberg calls "parameter robustness" (Weisberg, 2013, p. 159).

for studying the uncertainty of a model, and allocating it to the sources of uncertainty in its inputs. In the use I am considering here, it involves varying assumptions in order to determine the effect that these variations have on the results. For example, let us consider again an idealisation of no air resistance, justified by a negligibility assumption. If we have set the parameter representing air resistance in our model to $k = 0$, we might vary this by considering small but non-zero values of k (small relative to some natural scale determined by the problem). The aim is to ensure that the results we get don't depend sensitively on the air resistance being exactly zero, and simultaneously to test that the negligibility assumption (about the real system) holds in our model—i.e., that small values of k make only small changes to the results.

The result of this kind of investigation is what Frigg and Nguyen (2016) call a "key". By analogy with a map's key, this is a legend that tells the user how to interpret what they're seeing. It specifies how results from the model should be taken to relate to the world, covering issues of precision and realism: a key might specify that some precise number generated by the model should be taken as a prediction for the real system only to within some error-margin; or it might identify some element of the model as an artefact, not to be imputed to the target at all.

This trade-off is thought to prevent theorists from developing a single "best" model for a complex system (Levins, 1966; Teller, 2001; Weisberg, 2013, Ch. 9). The resulting prevalence of multiple models of a single system also has methodological implications—most straightforwardly, we cannot take disagreements between, e.g., Teller's two models of water as a sign that one of them must be rejected. Each can be useful for its purpose. Wimsatt (2007, p. 104) highlights that multiple idealised models can support the development of fuller theories, through the examination of results on which all models agree. This is a particularly useful technique in situations without underlying fundamental theory such as some areas of biology (Weisberg, 2013, p. 156).

As the above implies, criticising models is a complex business. As models have restricted domains, and specific purposes, the most natural way to critique a model is by examining its performance of its purpose within its domain. Performing poorly on other tasks, or in other domains, does not necessarily count against a model. It can do so if two models are being compared, and the one performs better on the shared purpose and has wider scope (either wider domain or the ability to fulfil multiple purposes). Put another way, models are not sensitive to counterexamples the way that fully general accounts are. Saying "here is a case that isn't like your model predicts" matters only if the case is in scope. Similarly, saying "your model says things are like so, but here is a case where they aren't" only matters if that feature of system in the model is intended to be imputed to the target. If the model's key identifies the feature as an artefact or says it should be imputed in some modified form, then the disagreement between the model's properties and the target's properties is irrelevant.

5 Normative models

Having reviewed these lessons from the philosophy of scientific modelling, I now turn to our main topic: normative models. My aim is to argue that formal epistemology fits the characteristics that define modelling, and therefore that the methodological considerations discussed above apply to formal epistemology too. But in order to do so, and in indeed in order for model-talk to go through, an important task remains: we need to provide an account of the main difference between inquiry in FE and the sciences: much of FE is normative. Our efforts are directed at what agents ought to do, rather than at explaining or predicting what they do. This section develops such an account. In addition to explaining what it means to say that something is a normative model applicable to real agents, I also want to vindicate our other common way of speaking: describing FE models as models of ideal rational agents.

To begin, however, let us note that normativity is not so foreign to science. Physiological models in medicine can be thought of as normative, representing how the body should be with actual deviations representing illness. Economic models of perfect competition might be taken by economists to specify how a market should work, with deviations representing barriers to be overcome, "imperfections" to be removed. Ecological models might represent an undisturbed ecosystem and thereby act as an evaluative standard for assessing the impact of alien species. Social choice models of voting procedures act as blueprints for the design of real voting mechanisms. Architectural models describe how buildings ought to be built.

Some of these involve a weaker sense of "normativity" than that familiar from ethics. But so long as normativity is understood as meaning "subject to judgements concerning oughts" then we can happily describe the above as normative for some sense of "ought". Nonetheless, current work in the philosophy of scientific models does not focus on these normative aspects, preferring representation as a topic of philosophical discussion. My purpose in highlighting these models at the start of this section is to point out that addressing normativity is not a concern peculiar to the application of model-talk to philosophy. Indeed, I will build my account of philosophical modelling by first considering models from outside of philosophy that play normative roles.

5.1 The architectural model

I will start with an example of one such normative model in science, as a guide to our thinking. Consider an architectural model of a block of flats. When architects first develop such a model, it serves mostly as a vehicle to communicate design ideas. The drawings are typically rough and impressionistic, representing high-level aesthetic ideas and establishing basic features of the building such as floorplan layout. As the building project advances, the model shifts to a more exploratory mode. Constraints from physics and engineering are incorporated, and a more familiar scientific use becomes dominant. Architects use the model to examine the implications of putting a staircase here, or opening that floor to

create more volume. Throughout the "design phase", it is a target-less model: it is not a representation of any existing building.

Later, the same model (now described by many complex drawings, outlining not just design and structural elements, but also services like plumbing and electricity) takes on a normative role for the construction team: it shows how the building ought to be built. There is a shift of audience over the design process, from client at the start to construction team at the end. For this latter audience, the target-less model becomes an instruction set for bringing a target into existence. There are two normative modes operant here: the model is an evaluative standard for the construction team's work, and it is action-guiding—skilled builders know how to translate the drawings into instructions. They build so as to bring into existence a building with properties as close as possible to those exemplified by this model. Once the building is complete the model becomes a familiar descriptive model, a representation of the new building (inevitably, an imperfect one due to deviations from the plan during construction).

This example gives us a handle on how normative modelling works. First, note that we have a movement back and forth between the model being descriptive and normative. This indicates that normative modelling is a way of using a model (rather than being a type of model). Which use it is put to depends on the purposes of the modeller/user and the intended audience. My first claim about normative models is thus: a normative model is any model that is put to normative purposes: evaluation, action-guidance, exploration of putative norms, and perhaps others.

Our philosophical models have similarly multifarious lives, a point that has been made in a different context by authors discussing the different "projects" of decision theory. Following Buchak (2013) we can distinguish four projects: construed normatively, decision theory can be used to evaluate or guide actions; construed explanatorily, it can be used to describe or interpret actions.[12]

The normative-evaluative use of decision theory involves analysing a decision situation facing an agent, and determining which actions are rational. The normative-action-guiding use involves deploying this process expressly in order to determine which act to undertake. The descriptive-explanatory use and the interpretive-explanatory uses are interested in real, rather than ideal, agents but they differ in their goals. Descriptive theorists are interested in describing observed patterns of behaviour – this is the empirical project of rational choice theory within economics. Interpretive theorists take real agents to be aiming at prescriptions of rationality but failing for various reasons. This theorist seeks to interpret the actions of the agent, as much as possible, as abiding by the rational theory of decision (this is often described as a "principle of charity").

The important point is that the *very same model* can be deployed in each of these projects (perhaps with different degrees of success). There need be no difference in the mathematical description of a decision theoretic model used in a normative-evaluative mode by philosophers interested in exploring the nature of

[12] I have relabeled these projects for convenience, taking inspiration from Thoma (2019).

rationality, and in a descriptive-explanatory mode by economists whose interest is in predicting choice behaviour. Note that the sequence that occurred in the architectural case (first target-less representational use, then normative) is inessential: a model can begin life as normative, and later be taken up for descriptive purposes.

More generally, it seems that we can fix some physical system as a reference point for evaluation generating norms ("do as Buddha did"), and thereby turn a descriptive model of that system into a normative model. This secures the movement from representational to normative use. In the other direction: any normative model will presumably be relevant to some actual system (e.g., set of people) that lies in the scope of those norms. I will call this the audience of a normative model; the architectural model's audience is the construction team. (If norms operate under an ought-implies-can principle, the real system could in principle realise the norms.) So, we can reinterpret any normative model as a (perhaps target-less) descriptive model of the relevant system where the norms are obeyed.

5.2 Idealisation

Our example "model" has normative constraints on the credibility relation, which correspond to what we take to be norms for partial belief. How can we fit these into the emerging account? Just as normative models are a use of a model, I will regard normative assumptions as a species of justification for idealisations. Recall that modellers might justify idealisations on the basis of negligibility, or because they delineate a domain, or as a heuristic device for making progress in early inquiry. Employing a normative justification for an idealisation is one way to put a model to normative use.

(This extends my error-theory of misused model-talk from philosophy to science: on this way of thinking, many economic models that are taken by their creators to be "positive" models are in fact normative, because they rely on assumptions like the transitivity of preference which the same economists justify by stating that it is a *norm for preference*.)

Consider our partial belief model, and the property of monotonicity. An economist might motivate this idealisation by any of Musgrave's three kinds of justification. The philosopher, by contrast, has a fourth option: whether or not it is approximately the case, or useful in simplifying analysis, partial belief ought to be monotonic over entailment. When philosophers and economists use "the same model," for normative and descriptive ends respectively, what they are doing is construing these conditions in different ways.[13]

Our normative models also contain artefacts, which is to be expected given their use of idealisations. Consider logical omniscience, which is also exemplified by the agents in our model of rational partial belief. While it is hard to

[13]However, note the parenthetical prior paragraph—some economists may want to insist (wrongly, on my view) that their ends are *descriptive* though they justify their idealisations normatively. The above refers to economists who justify these assumptions in other ways.

generalise about an entire discipline, my sense is that in practice logical omniscience is viewed as a mild embarrassment.[14] As I have set things up, it arises from monotonicity and the use of an objective logic to structure the Boolean algebra on which credibility is defined. These assumptions come before the representation theorem that gives us credences, and one of them (monotonicity) was assumed to be a norm of rationality. As a normative demand, however, it seems excessively strong (indeed it violates a widely held intuition that ought implies can). But shifting away from using an objective logic is a daunting task, however—models of bounded rationality are often complex (e.g., Garber, 1984). Many philosophers simply mark this property as non-normative—we do not want to continually criticise agents for their lack of logical omniscience—and continue to use the model with that built into the key. The same goes for our agents' abilities of instantaneous computation. The fact that Bayesians disregard these properties is evidence that they are employing a key, which is characteristic of modelling.

Thee example of logical omniscience raises an interesting complication, which will be discussed further below. Some of the idealisations (such as monotonicity in my example) in a philosophical model will be normatively justified, while others (such as completeness) will be justified in one of Musgrave's three ways. But these idealisations can interact to create the properties of the model. What do we say about properties that depend on both normative and non-normative idealisations? Can a property be anything but an artefact, if it emerges only because of a heuristic assumption? I will return to this in section 6.

5.3 Representation

In what sense is the architectural model representational, given that there is no real-world system that it designates?[15]

Philosophers have answered this question by making use of Goodman (1976) and Elgin's (1983; 2010) notion of representation-as. The idea is to separate out two parts of our ordinary notion of representation. Consider the example of a famous caricature of Churchill as a bulldog standing on Britain. This is a representation of Churchill, but in a sense it is also a representation of a bulldog (after all, it is a picture of a bulldog with a vaguely Churchillian face). We separate out these two notions by calling the first kind, involving denotation of a target, representation-of; and the second, involving the way in which the target is shown (also known as its secondary subject), representation-as. The formula is: an object X (the drawing) represents a target Y (Churchill) as something, Z (a bulldog).

The Z variable identifies the secondary subject; the kind of representation it is, or what it portrays. We will refer to these genres of representation as Z-representations (e.g., bulldog-representations). One aim of introducing Z-

[14] I report this as a sociological fact, to motivate for the existence of an implicit key. It is not universally true; some authors regard it as a serious problem—for a recent e.g., see (Bradley, 2017, Part IV).

[15] This subsection follows the explication in Frigg and Nguyen (2016, pp. 226–28).

representation is to show that there can be representation without reference to a target. A drawing of an orc is not a representation-of an orc, because there are no orcs. But the drawing does represent an orc in a sense, and we can now identify that sense by saying it is an orc-representation. As our formula says, Z-representations are objects (like drawings), and what fixes the genre Z is an interpretation. We can think of an interpretation as a function, mapping properties of the object to properties of Z. We associate properties of the caricature (particular lines, shading etc.) with properties of a bulldog (a certain stoutness, folded skin, etc.).

Interpretations allow us to talk about Z-representations as "having" Z-properties that, strictly speaking, they do not. (The drawing does not have four legs, it is a drawing.) With something like a caricature, certain properties are highlighted as particularly relevant and the intention is that we impute those properties to the target. Bulldogs are pugnacious, and the caricature highlights this with the stance of the Churchill-dog in the drawing, with the intention that we regard Churchill as pugnacious. This is the final part of representation-as: when there is a target, we can impute highlighted Z-properties to the target.

A number of popular accounts of scientific models agree that they utilise representation-as (Elgin, 2009; Frigg and Nguyen, 2016; Hughes, 1997). A model consists of an object and an interpretation $M = \langle X, I \rangle$. The object can be concrete like the V-shaped collection of plastic balls and rods, or abstract like the mathematical objects of the Bohr Model. The interpretation specifies what kind of representation the object is intended to be, for example by connecting the balls and rods with elements and chemical bonds. The model object has certain properties: the colour of the balls, the structure of the ball-and-link system, the type of links present, etc. These are mapped by the interpretation to various molecule-properties, such as elemental composition and bond-structure.

With these elements we can describe much of the everyday practice of "modelling". Consider again the Bohr Model. The (theoretical) model object is the orbital system, specified by a set of descriptions and equations.[16] This is interpreted as being an atom-representation. The descriptions and equations are part of the theory of quantum mechanics, and they provide guidelines for the manipulation of the model—where "manipulation" would here involve calculation. Various results can be derived about the model-system, which are spoken of in the language of atoms due to the interpretation.

The "models are representations-as" account is of particular use in explaining target-less models, like our architectural model. We can now say that it is a building-representation that is not a representation-of any building (before it is built). The development of the building design involves manipulation of the model, to explore and communicate various building-properties. In its normative phase, it remains a building-representation and this representational aspect of the model is necessary for it to fulfil its normative role. The way that it serves as an evaluative standard for the building is by being a building-representation;

[16] I will not discuss the details of how to account for what this theoretical object is. Frigg and Nguyen (2016, 2017) advocate a form of fictionalism about models.

exemplifying properties that the construction team's new building ought to have.

Our normative philosophical models have this representational aspect too. Our models of rationality are target-less; agent-representations that aren't intended to represent any real agents. These agent-representations are idealised in various ways so that the agents portrayed are dissimilar to real agents in various ways. Unlike scientific models however, some of these differences are regarded as normative. This account explains our language when we say that these are "models of ideal agents," and when we talk about those agents' beliefs and desires.

Much of our work in formal philosophy involves the manipulation of the model objects (the mathematical structures), in the form of deriving results and interpreting them in terms of the properties of agents. This is what we are doing when we prove a representation theorem, as discussed in section 2: we prove a theorem stating that our binary relation $\succeq$ can be represented by a probability measure, P. We extend the model's interpretation to cover this function and its properties: credences are the probabilistic partial beliefs of ideal agents. We can use the rich structure of probability theory to more easily manipulate this model object, and prove all manner of results—about the rationality constraints on partial belief in general, or about particular situations where we fill in the model description with additional details (say, about a decision an agent wants to make).

5.4 Purpose, scope and criticism

Scientific models are purpose-specific, with restricted domains of applicability. Given what has come before, it is hopefully now plausible to you that our mathematical frameworks in formal philosophy are models too, albeit with normative ingredients. So are they, too, purpose-specific and domain-restricted? In a weak sense of purpose-specificity, this might seem trivial. They are agential models of rationality, built to explore the rationality conditions on partial belief. That fills the basics of "purpose"—it tells us what the model is a model of, and what it is trying to do. But does this purpose also lead to modelling choices that restrict the model's usefulness in answering other questions? Are our models evaluated not on their truthfulness or truthlikeness, but instead on their adequacy for purpose? I think the answer must be yes (because I think we are modelling), but that this has not been sufficiently acknowledged by many philosophers working in FE. So here is one place where it matters that philosophers are (unknowingly) using the tools of modelling without acknowledging their limitations.

From the discussion above, we can discern one difference between the descriptive and normative cases. Normative models have an additional ingredient in the specification of their domain: the audience for normative guidance. They have one ingredient fewer, too: they lack a target of representation. The difference this makes is small, for the purposes of this section. The kind of inferences we draw from normative and descriptive models are different, and they apply to different objects (in the normative case, the audience; in the descriptive, the target). But in each case, we draw inferences from the model that are intended to

apply to some external object. The question is: what constraints this inferential process?

One answer comes from a consideration of purpose. Philosophers working with normative models put them to a number of different purposes. They might build a model to test a candidate norm—seeing what predictions emerge from a model employing it, and testing these against intuitions about what counts as rational. Such a modelling purpose will set implicit criteria for success: a good model for this purpose is one which can easily generate the test results in the sorts of cases. Or, they might aim to deploy norms for evaluative purposes (rather than testing whether they are in fact norms). Here, success will involve generating clear evaluative criteria. Finally, we may shift from evaluative to prescriptive normativity and seek to provide action-guidance. Success here will look quite different: action-guiding models need to be usable by those they provide guidance for, and this usability criterion may diverge significantly from the prior two.

Consider a Bayesian decision theory model. Agents in this model have probabilistic credences, they update their beliefs by conditioning on new evidence, and they make decisions by maximising subjective expected utility. These models do well on the first two purposes discussed above: they are simple to use to generate decisions in test cases like the Allais and Ellsberg scenarios, and establish clear criteria for rational belief, preference and decision. They are not very helpful for action-guidance, however. The process of eliciting any real person's attitudes and representing them as utilities and probabilities is onerous. As is often noted, Bayesianism demands too much of real agents. But note that this isn't a problem if the purpose is norm-testing or evaluation. It is only a problem if the modeller intends the model to be used for action-guidance.

6 Aside: what else could we be doing?

Some readers (though perhaps not those who have gotten this far!) might think: "Of course it is modelling, what else could it have been?" Why spill so much ink on the topic?

There are other methodologies that we might be using. Not all of science is modelling, and I don't think all of philosophy is either. Godfrey-Smith (2007) distinguishes the model-based "strategy" of science from an alternative, more direct, method of theorising. He contrasts two projects in late-twentieth century: Maynard-Smith and Szathmáry's *The Major Transitions in Evolution* (1995) and Leo Buss's *The Evolution of Individuality* (1987). The former is an exercise in model-based science, introducing many different models to isolate and discuss various causal mechanisms. The latter employs no models at all; instead, Buss examines actual organisms, in their actual circumstances. This work is close to the data, and involves studying real rather than fictional systems. It is synoptic, making progress by systematising knowledge (Godfrey-Smith compares Buss's work to Darwin's).

So, even if naturalistic philosophy recommends using scientific methods,

these needn't be modelling. This direct method, however, is not a good description of formal epistemology. We don't work from close attention to real agents, and not merely because we have normative aims. Consider Cassam's (2019) *Vices of the Mind*. This is epistemology done "from the ground up"—the theory of epistemic vices is built from a close examination of real cases involving real people, but the theory is put to normative ends. It is manifestly unlike the formal epistemology discussed here.

One difference, of course, it that it is not formal. But not all formal work is modelling either. Keefe (2000) is insistent that her supervaluationist work on vagueness is not intended as a model—an idealised, indirect representation of the linguistic phenomenon. She aims at a true description of the phenomenon of vague language (Keefe, 2000, Ch. 1). Accordingly, her methodology is different—it isn't idealised in the sense I have discussed here—and so are the success conditions for her work. As she notes, it is not open to Keefe to tell us to disregard certain parts of her mathematical framework as artefacts, or to isolate her account of vagueness from other accounts of linguistic functioning. Her work *is* open to refutation by counterexample, by design.

Modelling is a method rather than a goal. It therefore doesn't conflict with the traditional philosophical *project* of conceptual analysis, nor with ameliorative analysis, or Carnapian explication. As noted above, models are used to isolate mechanisms or concepts for particular study (Wimsatt, 2007, p. 15). Models can therefore support any of these projects by providing an isolated testing ground for an analysis of a concept or for a new concept. Similarly, multiple idealised models can support the development of fuller theories (which we might want not to be simplified or distorted), through the examination of results on which all models agree.

Given this, other readers might say: "modelling requires a conscious intentional stance to one's work, so if the people you describe don't *think* they're modelling, then they aren't."[17] I feel the force of this, but I think that some philosophers have been confused about their methods, or have lacked the conceptual machinery to realise that what they're doing is modelling. The aim of this paper is to furnish them with such machinery, in order that they can realise that they are modelling.

7 Methodology and inference

This brings us to our discussion of the methodology of normative modelling. As we just saw, criticisms of normative models must take heed of their purposes and their keys. To criticise a Bayesian model for properties that skilled users know to disregard—such as logical omniscience, or instant computation—is to misunderstand the methodology of modelling. That said, if a result depends in an important way on a property keyed as an artefact, it is similarly a methodological error to make use of that result—either imputing it descriptively to a

[17] Acknowledgement

target in the descriptive case, or making a normative inference relying upon it in the normative case.

This reflection allows us to formulate methodological constraints on the kinds of inferences we can and can't draw from normative models.

- Property X appears in our best account of rational partial belief. Therefore, agents are rationally required to have property X. (The "argument for probabilism from representation theorems" employs this move—e.g., in Maher (1993), and see (Hájek, 2008; Konek, 2019) for discussion.)

  One we replace the term "account" with "model" it becomes clear we need to be careful. In the descriptive case, realist inferences from discovered properties of the model to the target must be motivated with reference to their (in)dependence on idealising assumptions. Similarly, in the normative case, not all properties in the model are going to count as normative. Ignorance of a model's key makes it very difficult to cogently criticise. This leads to a methodological norm for modellers: be clear about what you regard as an artefact, and what you intend to be imputed to the target.

- Property X appears in your account. Property X is absurd, so your account is false. (Glymour's (1980) argument against Bayesianism repeatedly deploys this move.)

  As above, we now see that useful models may contain worrisome properties, which must not be imputed to the target. Sometimes we will need to avoid applying the model to cases where that property would do important work.

- Your account doesn't work in case Y. Y is a counterexample, so your account is false. (Very common, but e.g., the argument against imprecise probabilism in (Elga, 2010) has this form.)

  Models have a domain of applicability, so each "counterexample" must be checked against this domain. Objections irrelevant to the model's intended purpose have no bite. Instead, they motivate for a different model to be developed (perhaps to handle just those cases, or to expand the scope). Working out the boundaries of applicability for different philosophical models is a research area deserving of more attention.

As these moves are common, there is an important debate to be had about which bits of formal philosophy are modelling, and which are not. However, "it is just a model" should not be a get-out-of-jail-free card against objections (Keefe (2000, pp. 49–56) accuses some vagueness theorists of using it this way). This reinforces the need for clarity on the purpose and context guiding the modelling, and its key.

7.1 Securing normative inferences

I now want to return to the question raised at the end of the section (5.2) on idealisation. How can we know that our normative inferences are "secure" in

the face of their dependence on non-normative idealisations?

The problem concerning us here is that some of our normative conclusions depend necessarily on these assumptions. Consider Probabilism, the claim that one must have partial beliefs that are probabilistically representable. But, as we have seen, probabilistic representability requires that one's partial beliefs are complete and continuous.[18] In the scientific case, results which depend necessarily and sensitively on heuristic idealisations are generally regarded as artefacts and not taken to inform us about the target. Supposing for the moment that completeness and continuity are neither normative standards nor approximately true descriptions of partial belief, what does that mean for Probabilism? More generally, what can we say about when our normative inferences from models are secure?

Norm in, norm out

The easiest case is this: we construct a model in which all the idealisations employed are normatively justified. The agent-representation that results would differ from real agents only in ways that are normative. Then, it would be clear that the results generated by the model can be used to generate normative claims: norm in, norm out. But this is a limiting case that isn't that helpful— a purely normative model would be like a descriptively accurate scientific model. Its inferences would be secure, but it would not in truth be a "model". Models gain their usefulness from their ability to simplify through distortion and abstraction.

Approximate norms

Next, consider a model which employs some number of normative idealisations and also one non-normative idealisation, which is justified by a negligibility argument that it is a plausible justification in this case, perhaps bolstered by a domain restriction to the cases where it works best. (Perhaps we have a group of people who have nearly complete partial beliefs over some algebra of relevant prospects.) Now it seems we are on good ground. Our model is idealised, and we may wish to employ sensitivity analysis to determine whether the precise nature of the idealisation is generating any artefacts. But if the negligibility argument is sound, it provides good reason to think that we can find a model which captures the remaining properties of the system without introducing sensitive dependence on the idealised factor.

Now suppose we generate some results from working with the model—can we draw secure normative inferences for the model's audience? I think the answer is yes, but with a suitable understanding of what it is that models can achieve. Scientific models work well under these conditions, but their use involves an

[18] Properties of this sort show up in all of the representation theorems for credences that I am aware of. Without completeness, it is impossible to ensure that each proposition can be assigned a unique number in the way that credences do, and without continuity (or one of its any variants) one cannot get the rich structure of the real number line—in particular, the fact that when we have two real numbers, we can always find a third that lies between them.

acknowledgement of their limitations and fallibility. Models focus on what's most important, but in so doing they sacrifice precision. Their value is in capturing the main features of a system's behaviour, driven by the most important underlying factors. In complex systems this is remarkable, but it comes at the cost of precision. Philosophers often want to make definitive statements about norms, and may instead need to learn to present their results as "approximately normative," or as candidate norms that need to be confirmed by less idealised methods.

Formal epistemologists do not always do this. Recall that in our partial belief model, credibility needs to satisfy a long list of properties if we want to prove a Joyce/Villegas-style representation theorem: it must be monotonic, separable, transitive, complete and continuous. The chief normative result from this model that I considered was Probabilism, the thesis that if one's beliefs cannot be probabilistically represented then one is irrational. In many presentations, Probabilism has none of the qualifications or hedges that one sees in model-based science; nor is it accompanied by an analysis of the derivation's dependence on idealisations (e.g., Hájek, 2008; Joyce, 1998; Pettigrew, 2016). This is, I think, a mistake.

The first three of the assumptions about credilibility, I introduced as norms; the latter two are what decision theorists call "structural assumptions". That is to say that they are neither norms nor descriptive claims about real agents' partial beliefs, but instead are assumptions about the mathematical structure of the model object, introduced to make the analysis easier.[19] In our language, we can think of these as idealisations justified heuristically, or in some cases perhaps as domain-restrictions.

For example, economists are famously cavalier about completeness of preference, regarding it as delineating the scope of the problem they are concerned with.[20] (This is implausible, and much criticised (Joyce, 1998, pp. 98–103), but for the moment let us take the point to be: there is a practice of justifying structural assumptions in language that is recognisable to the philosophy of scientific modelling.) A parallel justification for the case of credibility would be to assert

[19] The term "structural assumption" appears to come from the measurement theory literature. Krantz et al. (1971) say: "Nonnecessary axioms are frequently referred to as *structural* because they limit the set of structures satisfying the axiom system to something less than the set determined by the representation theorem." (They are here referring to a more primitive representation theorem in which one establishes that the basic thing being measured—i.e. the attitude of partial belief—can be represented with an ordinal structure.) There are here using "structure" and "satisfy" in a set-theoretic and model-theoretic sense. The key point is that this reduction is meant to select a set of structures which are easy to work with. So in our case we reduce the set from those that are merely monotonic separable and transitive to the special subset which are also continuous and complete and thereby representable by a probability function.

[20] See for example (Arrow, 1966, p. 225), (Luce and Raiffa, 1957, p. 287), as well as a more modern example in (Gilboa, 2009, pp. 51–2). Gilboa describes completeness as normative, in the sense of being an injunction to the decision-maker: face-up to your decisions! But in the context of his presentation of preference as derived from choice, and of some choice in fact being made from a set of options, I think it is natural to describe his move as a domain-restriction.

that we are only modelling cases where an agent considers a limited number of prospects and does, as a matter of fact, make comparative judgements of likelihood about all pairs of prospects. This can succeed in making the idealisation innocuous, at the cost of reducing the model's scope (perhaps drastically).

A less restrictive justification would be to regard it as heuristic: a simplification for the time being. Something like this thought is present in the decision theory/FE literature in the form of the "coherent extendibility" thesis. This thesis states that there is nothing irrational about making no judgement about which of two prospects is more likely, but there is an important constraint contained in completeness nonetheless. A rational agent's credibility ranking must be such that it is possible to *extend* the relation to one that is complete, without violating any of the core rationality axioms (transitivity, monotonicity, separability) in the process (Jeffrey, 1992, p. 85) (Joyce, 1998, p. 103). This amounts to a complex key for interpreting the normative results of the model.

Complex dependency

More commonly, normative results (like Probabilism, and Savage's subjective expected utility theory) depend on a complex mixture of normative and non-normative assumptions. In this general case, I take the methodological import of the foregoing to be: with great power (idealisation) comes great responsibility (sensitivity and robustness analysis, humility in the face of model pluralism, careful attention to the model's purpose and domain). Modelling is a difficult business, and philosophers have thus far rarely exhibited the careful analysis required to extract secure inferences from their models.[21]

8 Conclusion

Once we accept that what we are doing is modelling, the implications for our philosophical practice are wide-reaching. One immediate impact is that it reveals a certain fruitlessness to the current debate between defenders of Precise Probabilism (the thesis we have already discussed) and Imprecise Probabilism (the weaker thesis that one's credibilities must be representable by a set of probabilities). What is at stake in that debate is a norm (i.e., the permissibility of averse to ambiguity), but much of the debate takes place at the level of model results, which are complex functions of normative and non-normative idealisations. If the models being compared employ different idealisations, and were built for different purposes, we now see that there is no way to compare them *tout court*.

[21] Some parts of the wider formal philosophy community already act in roughly the manner that I recommend here. In formal social epistemology, where much work consists of model building and application, there is a conscious effort to attend to good modelling methodology in a way that aligns with my recommendations here. For example see the discussion of (the lack of) stability analysis in Zollman (2010) by Rosenstock, O'Connor, and Brunner (2017) and Frey and Šešelja (2018). This work, however, is largely descriptive rather than normative.

Consider completeness again. One route to Imprecise Probabilism is to go the coherent extendibility route: the set of all permissible completions of a credibility relation generates a set of probability functions. For one purpose, the precision of a Precise model might be favoured and regarding completeness as a negligibility/domain idealisation may introduce no great difficulties. In another case, where completeness would obstruct her purpose, the modeller might use an Imprecise model instead. There is nothing odd about this state of affairs once we see that we are modelling. Note that while Precise and Imprecise Probabilism disagree over the norms of rationality, the class of Imprecise models includes the class of Precise models, and so Imprecisers are free to claim that certain contexts and purposes support the use of a Precise model. Precisers, on the other hand, are committed to the claim that only their models are acceptable. In the debate, their strategy must therefore be to block the permissibility of ever using an Imprecise model. In order to do so, they must either conclusively reject the norm at stake; or identify goals/purposes so universal that no model that does not accommodate them can succeed, and then show that no Imprecise model can do so.

The Precision debate itself is born of a sense that there must be a single, true normative account of partial belief. This is an admirable goal, but we must not mistake two models, idealised and distorted as they are, for candidates for such an account. One important lesson from scientific modelling is that a multiplicity of models is no bad thing! Each is likely to do best on its "home turf", and each will have different lessons for us about partial belief. Careful study of the characteristic problems for each model will help us to identify their home turfs and the boundaries of their domains of applicability. With these in hand we can turn to more important issues than fighting about whether Precise or Imprecise Probabilism is correct, such as looking at areas where neither does well. Here, a new model is needed.

Does this model pluralism commit us to antirealism? In truth, I am not sure. This may be a concern for philosophers more used to "direct" theorising about their domains. While I am not personally concerned if my conclusion is antirealist, there may be a careful path toward realism—it is certainly not the case that every modeller and philosopher of modelling is an antirealist. What they are, however, is very careful about elevating claims about the content of models to claims about the content of reality. Careful attention to "robust results" can reveal what is common between disagreeing models, which in turn may be candidates for realist inference. Careful de-idealisation is another route, though one which may prove intractable. In either case, formal epistemologists will benefit from attending to the philosophy of scientific modelling. Philosophers in general will find formal philosophy more accessible once they are equipped to evaluate its results as the outputs of a model.

References

Allais, M. (1953). "Le Comportement de l'Homme Rationnel Devant Le Risque: Critique Des Postulats et Axiomes de l'Ecole Americaine". In: *Econometrica* 21.4, pp. 503–546. ISSN: 0012-9682. DOI: 10.2307/1907921. JSTOR: 1907921.

Arrow, Kenneth J. (1966). "Exposition of the Theory of Choice under Uncertainty". In: *Synthese* 16, pp. 253–69.

Bovens, Luc and Stephan Hartmann (2003). *Bayesian Epistemology*. Oxford: Oxford University Press.

Bradley, Richard (2017). *Decision Theory with a Human Face*. Cambridge University Press. ISBN: 978-1-107-00321-7.

Buchak, Lara (2013). *Risk and Rationality*. Oxford University Press. ISBN: 978-0-19-175904-8.

Cartwright, Nancy (1989). *Nature's Capacities and Their Measurement*. Oxford University Press.

Cassam, Quassim (2019). *Vices of the Mind, From the Intellectual to the Political*. Oxford University Press.

Colyvan, Mark (2013). "Idealisations in Normative Models". In: *Synthese* 190.8. JSTOR: 41931906.

Elga, Adam (May 2010). "Subjective Probabilities Should Be Sharp". In: *Philosopher's Imprint* 10.5, pp. 1–11.

Elgin, Catherine Z. (1983). *With Reference to Reference*. Indianapolis and Cambridge: Hackett.

— (2009). "Exemplification, Idealization, and Understanding". In: *Fictions in Science: Essays on Idealization and Modeling*. Ed. by Mauricio Suarez. London: Routledge, pp. 77–90.

— (2010). "Telling Instances". In: *Beyond Mimesis and Nominalism: Representation in Art and Science*. Ed. by Roman Frigg and Matthew C. Hunter. New York: Springer, pp. 1–18.

Ellsberg, Daniel (1961). "Risk, Ambiguity, and the Savage Axioms". In: *Quarterly Journal of Economics* 75.4, pp. 643–669.

Eva, Benjamin and Stephan Hartmann (2019). "On the Origins of Old Evidence". In: *Australasian Journal of Philosophy* forthcoming in print, pp. 1–14.

Frey, Daniel and Dunja Šešelja (July 1, 2018). "What Is the Epistemic Function of Highly Idealized Agent-Based Models of Scientific Inquiry?" In: *Philosophy of the Social Sciences* 48.4, pp. 407–433. ISSN: 0048-3931. DOI: 10.1177/0048393118767085. URL: https://doi.org/10.1177/0048393118767085 (visited on 07/16/2018).

Frigg, Roman and Stephan Hartmann (2018). "Models in Science". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2018. Metaphysics Research Lab, Stanford University. URL: https://plato.stanford.edu/archives/sum2018/entries/models-science/ (visited on 07/04/2018).

Frigg, Roman and James Nguyen (2016). "The Fiction View of Models Reloaded". In: *Monist* 99.3, pp. 225–242. DOI: 10.1093/monist/onw002.
— (2017). "Models and Representation". In: *Springer Handbook of Model-Based Science*. Ed. by Lorenzo Magnani and Tommaso Bertolotti. Dordrecht, Heidelberg, London and New York: Springer, pp. 49–102.
Garber, Daniel (1984). "Old Evidence and Logical Omniscience in Bayesian Confirmation Theory". In: *Testing Scientific Theories*. Ed. by John Earman. University of Minnesota Press.
Giere, Ronald N (1988). *Explaining Science*. Science and Its Conceptual Foundations. Chicago: University of Chicago Press. URL: http://www.press.uchicago.edu/ucp/books/book/chicago/E/bo3622319.html (visited on 07/16/2018).
— (2004). "How Models Are Used to Represent Reality". In: *Philosophy of Science* 71, pp. 742–52.
Gilboa, Itzhak (2009). *Theory of Decision under Uncertainty*. Cambridge: Cambridge University Press.
Glymour, Clark N. (1980). *Theory and Evidence*. Princeton: Princeton University Press.
Godfrey-Smith, Peter (2006). "Theories and Models in Metaphysics". In: *The Harvard Review of Philosophy* 14.1, pp. 4–19. ISSN: 1062-6239. DOI: 10.5840/harvardreview20061411. URL: http://www.pdcnet.org/oom/service?url_ver=Z39.88-2004&rft_val_fmt=&rft.imuse_id=harvardreview_2006_0014_0001_0004_0019&svc_id=info:www.pdcnet.org/collection (visited on 07/04/2018).
— (Feb. 15, 2007). "The Strategy of Model-Based Science". In: *Biology & Philosophy* 21.5, pp. 725–740. ISSN: 0169-3867, 1572-8404. DOI: 10.1007/s10539-006-9054-6. URL: http://link.springer.com/10.1007/s10539-006-9054-6 (visited on 07/04/2018).
— (2012). "Metaphysics and the Philosophical Imagination". In: *Philosophical Studies* 160.1, pp. 97–113. JSTOR: 23262475.
Goodman, Nelson (1976). *Languages of Art*. Indianapolis and Cambridge: Hackett.
Hancox-Li, Leif (2017). "Idealization and Abstraction in Models of Injustice". In: *Hypatia* 32.2, pp. 329–46. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/hypa.12317 (visited on 01/06/2020).
Hájek, Alan (2008). "Arguments For, or Against, Probabilism?" In: *The British Journal for the Philosophy of Science* 59.4, pp. 793–819. JSTOR: 40072312.
Hughes, R. I. G. (1997). "Models and Representation". In: *Philosophy of Science* 64, S325–S336. ISSN: 0031-8248. JSTOR: 188414.
Jeffrey, Richard (1992). *Probability and the Art of Judgment*. Cambridge: Cambridge University Press.
Joyce, James M. (1998). "A Nonpragmatic Vindication of Probabilism". In: *Philosophy of Science* 65, pp. 575–603.
Joyce, James M (1999). *The Foundations of Causal Decision Theory*. Online (20. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511498497.

Keefe, Rosanna (2000). *Theories of Vagueness*. Cambridge, New York: Cambridge University Press.

Konek, Jason (2019). "Comparative Probabilities". In: *The Open Handbook of Formal Epistemology*. Ed. by Richard Pettigrew and Jonathan Weisberg. PhilPapers Foundation, pp. 267–348.

Krantz, David H., R. Duncan Luce, Patrick Suppes, and Amos Tversky (1971). *Additive and Polynomial Representations*. Vol. 1. Foundations of Measurement. Academic Press. Google Books: H6LiBQAAQBAJ.

Leitgeb, Hannes (2013). "Scientific Philosophy, Mathematical Philosophy, and All That". In: *Metaphilosophy* 44.3, pp. 267–75.

Levins, Richard (1966). "The Strategy of Model Building in Population Biology". In: *American Scientist* 54.4, pp. 421–431. ISSN: 0003-0996. JSTOR: 27836590.

Luce, R. Duncan and Howard Raiffa (1957). *Games and Decisions: Introduction and Critical Survey*. New York: Wiley.

Maher, Patrick T. (1993). *Betting on Theories*. Cambridge University Press.

Mills, Charles W (2005). ""Ideal Theory" as Ideology". In: *Hypatia* 20.3, pp. 165–185.

Mäki, Uskali (2009). "MISSing the World. Models as Isolations and Credible Surrogate Systems". In: *Erkenntnis* 70, pp. 29–43.

Musgrave, Alan (1981). "'Unreal Assumptions' in Economic Theory: The F-Twist Untwisted". In: *Kyklos* 34.3, pp. 377–87.

O'Neill, Onora (1987). "Abstraction, Idealization and Ideology in Ethics". In: *Moral Philosophy and Contemporary Problems*. Ed. by J.D.G. Evans. Cambridge: Cambridge University Press.

Parker, Wendy S. (2009). "Confirmation and Adequacy-for-Purpose in Climate Modelling". In: *Proceedings of the Aristotelian Society, Supplementary Volumes* 8. JSTOR: 20619137.

Paul, L.A. (2012). "Metaphysics as Modeling: The Handmaiden's Tale". In: *Philosophical Studies* 160.1, pp. 1–29. JSTOR: 23262471.

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

Rosenstock, Sarita, Cailin O'Connor, and Justin Brunner (2017). "In Epistemic Networks, Is Less Really More?" In: *Philosophy of Science* 84.2, pp. 234–52.

Suppes, Patrick (1969). "A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Science". In: *Studies in the Methodology and Foundations of Science*. Ed. by Patrick Suppes. Dordrecht: Reidel, pp. 10–23.

Teller, Paul (2001). "Twilight of the Perfect Model Model". In: *Erkenntnis* 55.3, pp. 393–415. JSTOR: 20013097.

Thoma, Johanna (2019). "Decision Theory". In: *The Open Handbook of Formal Epistemology*. Ed. by Richard Pettigrew and Jonathan Weisberg. PhilPapers Foundation.

Titelbaum, M. G. (2012). *Quitting Certainties: A Bayesian Framework Modeling Degrees of Belief*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199658305.001.0001.

Weisberg, Michael (2007a). "Three Kinds of Idealization". In: *The Journal of Philosophy* 104.12, pp. 639–659. ISSN: 0022-362X. JSTOR: 20620065.

— (2007b). "Who Is a Modeler?" In: *The British Journal for the Philosophy of Science* 58.2. JSTOR: 30115224.

— (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.

Williamson, Timothy (2006). "Must Do Better". In: *Truth and Realism*. Ed. by Patrick Greenough and Michael P. Lynch. Oxford : New York: Clarendon Press ; Oxford University Press, pp. 177–188. ISBN: 978-0-19-928888-5 978-0-19-928887-8.

— (2017). "Model-Building in Philosophy". In: *Philosophy's Future: The Problem of Philosophical Progress*. Ed. by Russell Blackford and Damien Broderick. Oxford: Wiley.

Wimsatt, William C. (2007). *Re-Engineering Philosophy for Limited Beings*. Cambridge, MA: Harvard University Press.

Zollman, Kevin J. S. (2010). "The Epistemic Benefit of Transient Diversity". In: *Erkenntnis* 72.1, pp. 17–35. ISSN: 0165-0106. JSTOR: 20642278.