

Modelling in normative ethics

Joe Roussos

May 2021

1 Introduction

This is a paper about the methodology of normative ethics. It is exploratory and intended to be provocative. What I am exploring is the notion that moral philosophy is, or ought to be, engaged in modelling.

What is a model? I shall spend some time on this, but here is a brief example to get us started. In order to study the population of fish in my local pond, I observe the fish feeding, breeding, and dying, for a few generations. I realise that the pond has a finite capacity for fish, due to their needs for space and competition for food. I observe that the population this week generally depends positively on the population last week, but that as the population reaches the capacity of the pond, crowding hampers population growth. In order to predict the population behaviour, I decide to use the following equation: $N_{t+1} = 4N_t(1 - N_t)$, where N is the number of fish in the pond divided by the carrying capacity, and t is a time index counting weeks.

In so doing, I am modelling the fish population. There are a few notable features of this model. I have represented some aspects of the fish population mathematically. In so doing, I have ignored many features of the real pond and fish, such the natural variation in fish size and reproduction. I am treating time as discrete, and counting only weeks. I have ignored certain factors which I know to influence the population level, such as fishing. I make no claims that this equation describes fish growth everywhere: the 4 is a parameter that I choose based on my local observations. These are characteristic features of modelling as a method.

Talk about “modelling” as a method of philosophical inquiry is increasingly prevalent, across various philosophy subfields. Williamson (2006, 2017) has argued repeatedly that modelling is an important method philosophy. In formal epistemology it is increasingly common for philosophers to describe their own practice as modelling (e.g., Bovens and Hartmann, 2003; Eva and Hartmann, 2019), and to advocate for it as a method (Leitgeb, 2013, p. 273).

In all these cases, the talk of modelling and model-building is an analogy with the common-place scientific practice of inquiry using idealised representations, which we have just met. I share with these writers a common goal: to illuminate and perhaps

improve philosophical methodology by thinking about existing commonalities with scientific methodology, and to advocate for the adoption of new methods. I shall do so for one part of moral philosophy. I take moral philosophy to be a broad heading which covers metaethics, descriptive or empirical ethics, deontic logic, and my focus here: first-order normative ethics. By this I mean the study of goodness and of right action which aims to provide norms and principles that govern moral behaviour. My claim is that normative ethics can, does, and should make use of idealised models like my model of the fish population in my pond.

Several things come out of this methodological claim. The first is a new way of understanding what is going on (and going wrong) in the theory/anti-theory debate in ethics. The second is a new way of understanding impossibility theorems in population ethics, and their bearing on ethics as a whole. Third, I show how the fact that we are modelling can be deployed in debates within ethics, using the debate over prioritarianism as an example. Finally, the fact that ethicists have (unknowingly) been modelling comes with certain methodological constraints for them. Most notably, models are not sensitive to counterexamples in the way that much of ethical theory is taken to be. If I am right, this requires a significant shift in how ethicists practice their craft.

2 Theory versus Anti-Theory

As a methodologist, I like to take as my starting point a disagreement between practitioners. So let us reflect on a bitter conflict in normative ethics: the “anti-theory” critique. Anti-theorists have long criticised a certain kind of ethical philosophising as misguided, doomed to fail, and besides the point. The target of their critique is given the name “theory”, sometimes “ethical theory” or “moral theory”. So what is a moral theory?

While the term is much contested, here are some characteristics attested to by both theorists and anti-theorists. A moral theory:

- provides a decision procedure for determining which actions are right or wrong (Chappell, unpublished; Fotion, 2014; Louden, 1992; Nussbaum, 2000; Timmons, 2012; Williams, 1981)
- is decidable; i.e., one can check whether any particular action or belief is correct according to the theory (Fotion, 2014; Nussbaum, 2000)
- is axiomatisable; i.e., can be stated by theorists in terms of a finite set of principles (Fotion, 2014; Louden, 1992).
- is general, or complete, or universal; i.e., it applies to all of, or a very wide range of, circumstances, people, action-types, and so forth (Chappell, unpublished; Fotion, 2014; Louden, 1990; Nussbaum, 2000)

- is or aspires to be the uniquely true theory. Theories compete; there is only one correct theory (Chappell, unpublished; Fotion, 2014) ¹

Some combination of the above is often taken to mean that there are no moral dilemmas (Louden, 1992, p. 97); given a complete description of the circumstances, the moral theory yields a single consistent verdict.² In addition to ruling actions right or wrong, moral theory is supposed to tell us what *makes* these actions right, thereby offering an explanation of their rightness (the same goes, *mutatis mutandi*, for goodmakers) (Timmons, 2012, p. 13).

Anti-theorists often take utilitarianism as the paradigm of a (problematic) theory; in its more ambitious forms, it exhibits all of these characteristics.³ But there is, of course, a spectrum of systematicity and ambition. Rawls’s *Theory of Justice*, which is the paradigm example of “ideal theory” in political philosophy, restricts itself to justice and indeed to articulating only some aspects of justice-as-fairness, leaving aside other normative questions.

This notion of “theory” is similar to and draws on the concept of a scientific theory. Scientific theories are tightly associated with laws of nature, regularities that are taken to hold very generally in a domain. Science begins with observations of particular facts, and proceeds by noticing certain patterns in the empirical phenomena. These patterns, sometimes called empirical laws, are one thing that science seeks to explain through theories. Theories aim to unify diverse phenomena by presenting the empirical uniformity they exhibit as the results of a common set of basic theoretical laws (Hempel, 1966, p. 75). Theories seek to explain that uniformity, offering explanations and understanding of the phenomena in question, and allowing for predictions via the laws. Theoretical laws involve the introduction of theoretical concepts, which go beyond what can directly be observed.

In the early twentieth century, philosophers began to analyse mathematical and scientific theories in formal languages (see Glymour, 1999, for a historical discussion). In one resulting tradition, theories came to be understood as sets of sentences in such a formal language. These sets are consistent, deductively closed, and (ideally) axiomatisable. It is from this tradition that we get the logical terminology used to characterise moral theories in the list above.⁴

¹Though I will focus on normative ethics, this conception of theory is taken up more broadly in moral philosophy. List and Valentini (2016, pp. 15–16), writing about political theory, use a definition of “theory” which includes a list very similar to the above.

²Occasionally it is also assumed that the theory is a set of sentences, which is consistent and deductively closed. This ensures that “theory” has the same meaning in ethics as it has in mathematical logic, and depending on how one formalises things, may make the no-moral-dilemmas aspect a consequence of this definition.

³E.g., Anscombe (1958) blames Sidgwick for bringing about the negative change that she detects in all English-language moral philosophy after him. Williams is another clear case, notably in (Williams, 1973, 1981).

⁴I am not claiming that the syntactic conception of theory is the *introduction* of the requirements of axiomatisation etc. Indeed, Aristotle’s philosophy of science has a central role for axioms, deduction, and consistency. I am here merely highlighting the *identification* of “theory” with such a set of

Why do “anti-theorists” object to theory, so described? There is an entire literature of arguments on this topic, so I will note but a few.

First, theory simplifies too much; it removes the nuance, complexity, and difficulty of moral reasoning. Bernard Williams is famous for this critique of utilitarianism. Reflecting on a pair of cases that he takes to be dilemmas, but which utilitarianism has ready answers for, he writes: “Not only does utilitarianism give these answers but, if the situations are essentially as described and there are no other special factors, it regards them, it seems to me, as *obviously* the right answers. But many of us would certainly wonder whether...that could possibly be the right answer at all; and...even one who came to think that perhaps that was the answer, might well wonder whether it was obviously the answer” (Williams, 1973, p. 99). The point is not that utilitarianism arrives at the wrong answer, but that it oversimplifies. This critique is not restricted to utilitarianism, either; McKeever and Ridge (2015) cite Raphael (1974) as deploying the same argumentative strategy against Kantianism.

Relatedly, theory is said to be too abstract, and too coarse, to deliver usefully precise recommendations in real situations. Any theory will “run out of gas before it reaches the level of concrete decision making required by practical ethics” (Arras, 2010, S3.3). Arras claims that theory will often leave too many options on the table, not because they are in truth morally equivalent but because theories are incomplete in an important sense.

Next, there are too many moral theories. For Baier, the very proliferation of theories shows that they are unlikely to succeed. This is because the ethical domain is simply too diverse to support successful theorising; it is not unified in the way that is required for theory to succeed. “Where do we have genuine and useful theories? Primarily in the sciences—but there we find a plurality of them primarily over time, rather than at a time. We certainly do not find some engineers building bridges or spaceships by application of one theory, while others at the same time are applying another different theory” (Baier, 1989, pp. 33–34). For Arras, this glut of contrary theories without an obvious choice between them makes theory useless for practical ethics, in the sense of providing a useful guide to practice (Arras, 2010, S3.1).

Finally, theories require “principles which are definite in meaning in order for them to play their role in the deduction of particular moral judgements. On the other hand, the norms of actual moral practices are vague in order to permit context to play a role in determining their application” (Clarke, 1987, p. 238). As Baier argues, a seemingly clear norm such as “don’t kill” “brings with it a very rich cultural baggage, if it is to have any content at all. Either it is a purely formal moral code, not yet prohibiting or enjoining anything, or else the form gets its determinate filling, in which case we are committed not merely to these ‘negative’ rules but to the rules of background institutions and ways of life that supply the determinate content to these prohibitions.” Theories, with their focus on the norms alone, are thus unable to stand in the required justificatory relation to actual moral practices (Baier, 1985, 273–74, quoted in Clarke).

sentences.

There may be something right to the anti-theory critique of certain overly-ambitious bits of ethical theory. But there is something deeply wrong with how the space of methodological options is characterised in this debate. Anti-theorists often seem to take the options to be “theory” (bad), or a form of very granular analysis that makes no attempt at generality or systematicity (good). In so doing, they neglect a middle-ground of partial systematisation, making use of intermediate principles with application to limited but still substantial domains. In science, models cover this middle-ground. One goal of this paper is to explore that middle-ground and advocate for its use.

3 Modelling in ethics

In this section I present an overview of modelling in ethics.⁵ I will not give a precise characterisation of what counts as a model in normative ethics, because no such characterisation exists in science either. I see many different models in ethics, of different kinds. Some are normative models of agents, and operate in a way that is similar to scientific models in that they represent some parts of the natural world. Others are models of goodness, and operate more like abstract models in theoretical physics, representing something that isn’t a concrete part of the natural world.

Let us begin by noting that some of the criticisms levelled by anti-theorists at moral theory also apply to scientific theories. Their abstraction and generality makes them hard to work with when explaining specific phenomena, or making predictions about a certain system. Scientific theories, and in particular theoretical laws, present an overly simplified picture of things, such that if one were to make observations “in the wild” one would *not* observe the behaviour predicted by the laws of nature as stated in, for example, Newton’s mechanics (Cartwright, 1983). But these complaints have no bite against scientific theories—certainly no one would suggest abandoning theorising on the basis of them.

One reason for this is that in science an important mediating role between theory and world is played by *models* (Morgan and Morrison, 1999). Models are representational structures, which make use of bits of theory (or sometimes different theories), diagrams, knowledge about instruments, approximation schemes, and other tools that are not part of any theory (Cartwright, 1999). Models are characterised by their use of idealisation—the distortion or omission of properties of the system they represent (Frigg and Hartmann, 2018; Weisberg, 2007a). Models are purpose-specific tools in ways that theories are not. The idealisations used to facilitate the model-based analysis must be justified somehow, and this justification is usually relative to specific goals of inquiry. The goals and idealisations together set a domain of application for the model—outside of this it should not be expected to work well (Teller, 2001; Weisberg, 2007b).⁶

Here is a high-school physics example. Newton’s theory of mechanics consists

⁵This section is somewhat anthropological. Having observed and interacted with ethicists, I offer a characterisation of their practice as modelling. This is a characterisation that would be more naturally made by themselves, once modelling talk enters their methodological lexicon.

roughly of his laws along with some core notions like “centre of mass” and “reference frame”, and a value for the parameter representing Earth’s gravitational attraction. A Newtonian prediction of what happens in a particular situation, even a highly stylised one, required a model. Consider a canon being fired: a model of this includes an idealised representation of the ball and the force of the blast, along with the canon’s initial orientation and elevation. The modeller might idealise the shape of the ball as a perfect sphere and assume there is no air resistance. They might make essential use of diagrams like the “free-body diagram”. Such a model is *required* for prediction or explanation: theories are too abstract to do that work for any particular circumstance.

In addition to mediating between theory and reality, Frigg and Hartmann (2018) discuss several ways that models and theories may interact. Models may be a means to explore a theory, or to complement one. These uses may occur when the theory is very complicated and difficult to apply in full. Or perhaps the theory leaves open certain questions, which a model fills in for particular cases. Models can also make quantitative what was only qualitative in the theory. Some models exist entirely without theory: in poorly theorised domains, model-building can be the first step towards building a theory (Wimsatt, 2007, p. 104).

I think this all happens in ethics too, though it is rarely explicitly recognised. The picture I have in mind is this.⁷ Like scientists, moral philosophers begin with a set of “data”: observations of moral life, and our moral judgements. Anti-theorists are right that, like many natural domains, the ethical domain is extremely complex and we have only partial information about it. Ethicists discern certain patterns amongst these, which they investigate, seeking eventually to systematise them in a moral theory. There may be some empirical regularities (e.g., common judgements, apparent norms), which we aim to explain by the introduction of theoretical concepts (e.g., precisified notions of duty, or welfare). But the domain is complex, patterns are hard to discern, and the data often seem contradictory, and so we cannot read off moral laws from the data.⁸ Ethical theorising therefore proceeds by modelling: restricting oneself to a limited domain (such as distributive questions for social planners, or justice-as-fairness), making distorting idealisations in order to simplify and facilitate systematisation.

(I hope it is obvious that ethicists regularly idealise. Sometimes this takes the form

⁶“Model” is one of those unhelpful terms that is used to mean many different things, so I want to flag 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—i.e., all models are set theoretic structures. I am not using “model” in this sense. My analysis of “model” draws on the philosophy of science tradition that includes Cartwright (1983, 1989) and Giere (1988, 2004), as well as many of those cited in this section.

⁷This is intended as a sort of rational reconstruction, rather than a description of the actual history of ethics.

⁸All theorists must accept this, I think, as part of their explanation for why so few are adherents to their theory.

of assuming that “all other factors are equal”, or bracketing questions about how a case arose [who tied those people to a railway?] which would be important in any real case. It may involve a consideration of only “laundered and fully informed preferences”. I will discuss a few idealisations in more detail in sections 4, 5 and 6.)

One aim of ethical theorising has been the pursuit of moral laws to parallel laws of nature. This is the project of moral theory. A notable fact about the ethics landscape is that we have many candidate theories competing to explain the same domain. The familiar examples being utilitarianism, Kantian deontology, and virtue ethics. This situation of competition occurs also in science, typically at times of theoretical uncertainty and change. Each theory proposes its own “laws”, and the dispute between them is in part about which laws fit the data best. One of my claims in this paper is that some criticisms of these moral theories show a need for models as mediators. Philosophers of science will not be surprised to hear from bioethicists like Arras (2010) that trying to directly apply theories to particular real world cases was unsuccessful. Each theory’s work can be fruitfully augmented with models. As in science, we should also expect to find models operating without theories.

What do models in ethics look like? There are some clear instances which can start us off. They are found in formal work, incorporating explicit idealisation, performed by philosophers with mathematical and scientific training that leads them to speak in terms of models and modelling. For example, McCarthy, Mikkola, and Thomas (2019) present a social aggregation theorem. This works very much like the scientific examples I have given: mathematical structures are introduced which represent things in the domain being studied: populations, values, preferences. Ethical principles are likewise represented in the model, as mathematical constraints or relations or properties of objects. The purpose of studying the mathematical structure is to learn about something else: how to think about the relation between individual and social welfare.

Beyond such clear examples, I think modelling is widespread in formal and semi-formal ethics, including work on value theory, welfare economics, distributive theory, population ethics, and moral uncertainty. It is easiest to spot in situations where the work is formal—i.e., takes place in a mathematical or logical vocabulary—because such approaches encourage precision about what is being assumed and idealised. (This is not to say that all formal work is upfront about its assumptions.) I think that most formal work in ethics involves modelling, but modelling does not require formalism! Nor does modelling have any necessary connection to work in the orbit of utilitarianism.

Some models in ethics are intended to be “predictive”—which I interpret here as meaning rendering the correct judgement about a case. The model is fed a scenario (e.g., described in a vignette about people tied to train track) and it delivers a conclusion which is then tested against “the data”—here, almost always our considered moral judgements.⁹

⁹In science, care is taken to separate out which data are used for testing the model. While building a model, the modeller may make use of certain data to calibrate it—ensure it gives the right answers, by adjusting certain parameters. Once the model is ready, it is tested against different data from that used to calibrate it. Success against this new data is taken as confirming the model’s usefulness, while

Other models are tested by the quality of the explanations they offer. These focus on rightmaking or goodmaking features. Here, having the right implications is not satisfactory; we want the right reasons for those implications. This too is common in science, where there is a large literature on different forms of explanation, and its link to understanding. What is missing in science, and *sui generis* in ethics, is the link to justification and action. But it is worth noting that in the scientific case the goals of explanation and prediction can come apart, with some models faring well on one and poorly on the other. Perhaps in ethics we shall find models which excel at “getting the answer right” but cannot give us a compelling story about why it is the right answer.

The purpose-driven nature of modelling means that we encounter multiple, disagreeing scientific 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). Each is highly successful at its purpose, i.e., prediction of the relevant kind of behaviour. Moreover, they contradict one another: one says that water has particles, the other says it does not. The lesson is that neither should be thought to provide a definite characterisation of water, and our understanding of water is enhanced by having both available. I think something analogous happens in ethics.

You can see now why it matters that we are modelling. Recognising that moral theorising involves modelling requires a re-conception of what it means to succeed. As the pat phrase goes “all models are wrong, but some are useful”. Philosophers are not used to thinking their claims are certainly wrong, which leads to certain methodological habits which we need to unlearn. 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. More subtly, models often 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. A common method for identifying such effects is sensitivity analysis. The modeller varies idealisations—introducing minor air resistance, considering non-spherical canon balls—to test whether their interesting results are mere artefacts of the unrealistic assumptions they made in order to make the analysis tractable.

success against the data used to calibrate the model is taken to be trivial. I am not sure whether there is a parallel to this in the normative ethics case.

4 Anti-theory redux

The theory/anti-theory debate is a dispute about whether ethics should be in the business of building theories, where that term is taken from or at least closely analogous with scientific theory. In this debate, theories are assumed to have the structure outlined above including, crucially, universality in the scope of laws (or perhaps definitions, in the case of value theory), and with entirely general domains of application. I propose that, insofar as the anti-theory critique does well, it often motivates instead for modelling.

I will offer two interpretations of the anti-theory debate in terms of the philosophy of science concepts introduced above. Which interpretation is better will depend in part on what ethicists take themselves to be doing. They need not compete—the first interpretation might better fit some parts of ethical theorising, and the second, others.

First interpretation: We have ethical theories, or partial theories, and these are precisely the targets of the anti-theory debate: utilitarianism, Kantian deontology, neo-Aristotelian virtue ethics, and so forth. These are abstract and distant from the phenomena that they ultimately describe. They contain laws, which opponents point out often generate the incorrect answer if applied naïvely. Practitioners reply that applying these laws requires skill, which we now interpret as the familiar claim that one must learn how to use models, approximation techniques, and various instruments, to connect these theories with reality.

Second interpretation: We do not have (successful) ethical theories. We are adrift in a complex and confusing domain, and our attempts at systematic investigation should be thought of as modelling in the absence of theory, modelling which hopes to develop a theory. What is often presented as a law is more like a model-bound regularity whose true domain of application is under investigation.

We begin with the first: there are theories, but we need models to connect them with reality. I think it will be helpful to begin with laws, and how they are thought of by philosophers of science in the modelling tradition I am presenting. To the extent that science involves genuine laws of nature (exceptionless generalities) they are thought to be the laws of physics. But Nancy Cartwright has forcefully challenged the notion that, even there, we have such laws. In *How the Laws of Physics Lie*, Cartwright argued that many laws of nature are literally false—what they tell us is not what happens. Those which are true are *ceteris paribus* laws, and apply only under abstracted and idealised conditions that are rarely realised in nature (Cartwright, 1983). Laws which aren't presented as *ceteris paribus* may have implicit conditions attached. Laws are thus true and serve an important explanatory function, but much of their work is done through models. The idealisations in these models serve in part to create situations in which the laws can literally apply. These models don't correspond to exact reality, yet they allow the theory to do its work (Cartwright, 1989).

This interpretation offers us a way of understanding and responding to the first anti-theory argument I discussed above: theory simplifies too much. We now see that theories in science are themselves highly abstract and distant from the empirical phenomena

they purport to explain. Anti-theorists have acknowledged this: Williams argued that the crucial difference is that in science the theories answer to the truth, which allows for scientific theories to be successfully general despite their abstraction. Ethics, for Williams, was inherently local, and he leaned towards non-cognitivism (Fotion, 2014, pp. 55–56).

Setting aside the metaethical question, I think Williams neglected the role of models in science and their potential mediating role in ethics. Consider Williams’s objections that utilitarianism doesn’t reflect the operations that real agents would carry out: the one-thought-too-many objection or the complaint that utilitarianism offers easy verdicts to difficult questions. If we consider a utilitarian model of one of his cases, we should expect that not all of it is intended to correspond to reality. There is nothing methodologically suspect, to the modeller’s eye, in claiming that this model is intended to generate successful predictions (i.e., render the correct verdict on the case) but *not* to represent its difficulty. The operations required of the modeller to produce the result may have nothing to do with the cognitive processes by which that verdict would be arrived at by an actual agent.

Baier’s argument that seeming laws like “don’t kill” are woven in to a cultural fabric which provides interpretations, exceptions and specifications now seems like nothing more than Cartwright’s analysis of “how the laws of physics lie”. Cartwright’s claim in the scientific case was that a careful understanding of laws as *ceteris paribus* generalisations, coupled with close attention to causation, would allow laws to come out as true, and to play an explanatory role in science. (In the next section I will return to the role causation plays for Cartwright and what might be analogous in ethics.)

On this interpretation, the bioethics critique of “high theory” as being unhelpful to that project is correct, unsurprising, and no real challenge to the theories themselves. Whether one works “top-down” or “bottom-up” in bioethics may depend on how successful one takes moral theories to be *as theories*. If one is independently inclined to think that none of the major moral theories is much good, then one would naturally want to work “bottom-up”, modelling in the absence of theory—in a way recognisable to any philosopher of biology. If, on other other hand, one thinks that a certain theory is broadly correct, one is more likely to work “top-down”, using a model to mediate between the theory and real-world cases. All theories need mediating models of this sort.

The second interpretation does better against the other anti-theory arguments I discussed above. On this interpretation we have no theory, and what we call a theory (e.g., utilitarianism) is better understood as a model.

Consider the claim that our moral lives contain irremovable moral conflicts or dilemmas, and that “theories” must therefore be false. This is less concerning if we substitute theories for models. Models are false, but hope to be useful. The lack of dilemmas in the model could be a form of idealisation justified as a simplification that is made in order to facilitate analysis. Perhaps the usefulness of the abstraction is then in illuminating the connections between various concepts, or seeing how they work together to generate conclusions. Or perhaps it could be justified as a domain restriction: this is

simply a model of cases without moral dilemmas. In those cases, it might be claimed that the model generates the right result.

This no-theory interpretation can also answer Baier's definiteness worry. She claims that the nature of "theories" is such that the norms which feature in them have properties that our actual moral norms do not have. The modeller can here respond that models precisify observed norms into principles for particular purposes, in limited contexts, without claiming that the representation of the actual moral norm in the model is *identical to* or *underlies* that norm. The precision facilitates a certain kind of analysis.

Finally, recall that the very proliferation of contradictory moral theories was taken by Baier as evidence that the project of "theory" cannot, or is at least very unlikely to, succeed. This objection seems tailor-made for the modelling response. It is one of the distinctive features of modelling that we find a proliferation of models which overlap and even contradict one another and yet, in their patchwork fashion, contribute to an overall understanding of their common domain. On this view our different ethical models might be like Teller's two models of water. They fare best in particular areas, explicitly conflict on some questions, and cannot be complete descriptions.

This brings us back to two important features of models discussed above. First, they are purpose-specific, and thus have restricted domains of application. Second, this means that they are not sensitive to counterexamples in the way that fully general theories are.

What could it mean to say that utilitarianism, say, has a particular purpose or restricted domain? These domains could be types of question, as I will discuss below for prioritarianism, or something as general as Nozick's "push" and "pull" factors for morality (Nozick, 1981). However they are spelled out, the result will be that certain questions simply aren't meant to be addressed by the model. This may (and probably will!) seem unsatisfactory to the ethicist used to debate by counterexample. If a theory is doing poorly in the general case, why trust it in a limited domain? We are rightly suspicious of a theory that says you can sometimes torture children and should feel uncomfortable about using it in non-child-torturing situations!¹⁰

There are three parts to the modeller's reply. The first is simply to insist that we are working in a complex, contradictory domain. *We do not have theories*, in the strong sense discussed above, though we are engaged in theorising. All of the available models are limited, and face "counterexamples", be they child torturing or Nazis at the door. The second part of the response is to articulate, in a non ad hoc way, a domain restriction. Some contrary piece of data only *fails* to be a counterexample if it is genuinely outside of the model's purpose. "It is just a model" cannot be a Get Out of Jail Free card, it is a description of a careful and principled methodological approach. Third, the modeller notes that we can still have conflicts between models and judge one better than the other. Consider one model, with a particular purpose and associated domain, outside of which it advocates for torturing children. Now consider a second model which has a wider domain—it can answer the same questions as the first model,

¹⁰Acknowledgement.

and more. On the common domain, the second model does as well as the first. The second model's wider domain includes the child-torturing cases, and it does not deliver the same incorrect result. In that case, the first model is clearly worse than the second. Worse for what? Well, for all purposes the two models have in common, and for general use—having wider scope is a theoretical virtue.

4.1 Moral particularism.

The foregoing discussion also gives us a way to think about moral particularism. Particularism is something like the view that not only is ethical theory impossible, but there is no middle-ground whatsoever. We must confront particular cases in all their granularity, rather than attempt any systematisation (e.g., Dancy, 2017). Modelling offers us a way to access precisely the middle-ground that particularists deny, however. It makes no claim to universality, or general application. Models can be local, they can synthesise only some of the available data. Importantly, they need not be axiomatisable, or decidable, or even formal. They are the tools of scientists engaged in the sort of ground-up work particularists seem to want us to engage in, but they achieve more in the way of generality than they take to be possible.

Writing about normative models in decision theory, Michael Titelbaum comments thus on particularism: “The normative modeler proceeds piecemeal, trying to solve local problems and gradually extend the boundaries of normative knowledge. (In this she is much like the working scientist.) The modeler does not fully yield to the particularist’s insistence on treating each case on its own terms, but neither does she assume that the normative is a single, systematizable domain” (Titelbaum, forthcoming, p. 16).

The hard-line particularist will reply that this is doomed to fail because it assumes that moral considerations function the same way across circumstances. For example, consider Dancy’s reasons holism, under which a feature of a situation, like the fact that someone is lying, can have a different moral *valence* across cases. (Dancy, 2017, S3)

This brings us back to Cartwright on laws of nature. Recall that to Cartwright, laws are literally false in much the way that particularists claims that the maxims in ethical theories are false. Cartwright argued that using laws requires the postulation of capacities, which *act in the same way in all circumstances*, despite the apparent falseness of the lawlike statements of science.

The logic that uses what happens in ideal circumstances to explain what happens in real ones is the logic of tendencies or capacities. What is an ideal situation for studying a particular factor? It is a situation in which all other “disturbing” factors are missing. And what is special about that? ...This tells you something about what will happen in very different, mixed circumstances—but *only if you assume that the factor has a fixed capacity that it carries with it from situation to situation*. (Cartwright, 1989, 190f, my emphasis) quoted in (Reutlinger et al., 2019)

In science, these capacities are causal powers which clearly won’t do for ethics.

Under our analogy, the reasons holist denies that morality has anything analogous to nature's capacities. The modeller, or aspirant theorist, thinks that it does, that there are goodmakers and rightmakers which have the same action across situations. This clarifies what the debate is about. The mere fact that moral laws don't straightforwardly apply in observed cases is not an argument in favour of particularism, as it does not undermine the existence of these capacities. What particularists need is arguments against the existence of such goodmakers and rightmakers, as revealed by a careful Cartwrightian analysis.

5 Impossibility theorems in population ethics

In this section I want to observe some idealisations in population ethics, and comment on how my modelling view might illuminate results in that field.

Population ethics is the study of ethical problems concerning populations—groups lives, people living for a given time with a given level of welfare. It is concerned with actions which affect how many people will live at a future time, and which people they will be. Amongst other things, it seeks a population axiology; that is, an ordering of populations with regards to their (intrinsic) goodness (Arrhenius, forthcoming). It often proceeds by thinking about which of two possible populations is better. The standpoint in population axiology is not one of considering action, for example bringing each population into being, but rather a judgement of their relative goodness. Following Derek Parfit's presentation of his "mere addition paradox", it has been recognised that there are significant difficulties in formulating such an ordering (Parfit, 1984, Ch.19).

One popular strand of population ethics focuses entirely on welfare. (Conceived, very roughly, as how well a person's life is going; how good it is for them.) It is in this context that these paradoxes and associated impossibility results arise. But they are not to be regarded as merely a problem for welfarists, claims Gustaf Arrhenius:

Since we can assume that other values and considerations are not decisive for the choice between the populations above, as we shall show below, this is not true. Hence, paradoxes like the above are a problem for all moral theories which hold that welfare at least matters when all other things are equal. Since, arguably, any reasonable moral theory has to take this aspect into account when determining the normative status of actions, the study of population ethics is of general import for moral theory. (Arrhenius, forthcoming, 5, emphasis mine)

As Arrhenius puts it, this focus on welfare is not because other considerations—such as fairness, liberty, and virtuousness—do not matter. They may well figure in the ranking of populations. But the population ethicist assumes "that welfare at least matters when all other things are equal". This is a clear idealisation—an omission of these other factors, on the grounds that they are being assumed to be equally balanced in the weighing of considerations. Put another way, it is a *ceteris paribus* clause. As

we've just seen, these play a crucial role in the strand of philosophy of science I am drawing on.

How are we to interpret the results of population ethics, given this idealisation? One of Arrhenius's contributions is to present precise theorems showing the impossibility of satisfying various conditions which are taken to be necessary features of an adequate population axiology. He proceeds by first introducing such a condition informally, on the basis of intuitive responses to cases. For example, avoiding the Repugnant Conclusion is one condition of adequacy. This is the result that, for a possible population of many high-quality lives, there is some much larger population of people living lives barely worth living, which is ranked *better* than the former by the population axiology. In general, Arrhenius's method is to first formulate an adequacy condition in words, on the basis of the relevant intuition-eliciting case or reflection, and then to introduce an exact formulation which employs mathematical representations.

Here is an example of a condition which is part of the precisification of avoiding the Repugnant Conclusion.

Quality: There is a perfectly equal population with very high positive welfare which is at least as good as any population with very low positive welfare, other things being equal.

Quality (exact formulation): There are two positive welfare ranges $R(u, v)$ and $R(1, y)$, $u > y$, and a population size $n > 0$, such that if $W_z \subset R(u, v)$, $A \subset W_z$, $N(A) = n$, and $B \subset R(1, y)$, then A is at least as good as B , other things being equal. (Arrhenius, forthcoming, p. 304)

This seems to me to clearly be a model.¹¹In addition to the basic feature that Arrhenius is employing mathematical representations to make his arguments precise, I note two other characteristic features of models: (1) "structural" assumptions are introduced to facilitate the mathematical representation, and (2) idealisations are introduced to simplify the analysis.

We have already discussed one idealisation involved: the focus on welfare. As an example of a structural choice, Arrhenius uses sets to represent welfare levels and he assumes that the set of welfare levels is fine-grained, in the following sense (Arrhenius, forthcoming, p. 299):

Finite Fine-grainedness: There exists a finite sequence of slight welfare differences between any two welfare levels.

So, what are we to make of the fact that this work involves modelling? Arrhenius presents his work as illuminating something about the structure of value, or of our intuitions about value. He is careful in his conclusions:

¹¹Though we needn't go into the details, here is a brief explanation: A welfare level (e.g., W_u) is a set containing populations (e.g., A) of equal welfare, where population itself is a set of lives. The number of lives in a population is denoted $N(A)$. A range of welfare (e.g., $R(u, v)$) is a collection of ordered welfare levels, represented by its top and bottom points. So $R(u, v)$ is the set of welfare levels starting with W_u , the lowest ranked level in the set, and ending with W_v , the highest. The special symbol "1" is reserved for the welfare level that is just above the level at which life is not worth living.

If the evaluations above stand up to scrutiny, that is, if we find it impossible to give up any one of them, then our considered moral beliefs are mutually inconsistent. And if consistency with considered intuitions is a necessary condition for a moral theory to be justified, we seem to be forced to conclude that there is no such theory which can be justified. In other words, paradoxes of the above kind might challenge some of our deepest beliefs about moral justification and the meaningfulness of moral theories. (Arrhenius, forthcoming, p. 4)

So, if these are the data, and fitting all the data is a requirement for a theory, then there is no moral theory (Arrhenius, 2000, forthcoming). But here I would make a friendly amendment: If these are the data, and fitting all the data is a requirement, *and this model—with its idealisations and structural assumptions—tells us something general about value*, then there is no moral theory. The italicised addition is crucial.

As Cartwright shows, when other things are not equal, modelling is much more difficult than in the ideal case. In Cartwright's picture, the movement out of the idealised case is licensed by nature's capacities acting in fixed ways from situation to situation. Modellers must engage in careful work to get their results to apply in messy real situations, either by sensitivity analysis, de-idealising the model, or presenting their results with explicit provisos linking them to the assumptions under which they were generated.

Now, let us suppose that Arrhenius's results *do* show that we can have no consistent *theory* of value, which captures all of this data. The population ethicist need not despair. There are many domains of science in which we have no overarching theory, or where we know that two successful models of sub-domains cannot be unified in a consistent manner. Fundamental physics is just such a case, where quantum field theory and general relativity, each highly successful in its domain, cannot currently be made consistent.

The modelling strategy is to go local, and construct models which capture some of the data, in some circumstances. As modelling is purpose-driven, this may require population ethics to become more applied. By responding to real-world problems, population ethicists may be able to reject an assumption, or to prioritise which of the conditions of adequacy are most important. This sort of purpose-driven prioritisation would then motivate the construction of a more local model of a population axiology—one which is known to be incomplete, but which can still be useful.¹²

6 Models of prioritarianism

In distributive theory, philosophers discuss the plausibility of distributive principles with respect to short vignettes presenting cases. This is another clear case in which

¹²This is similar to the approach taken by Budolfson and Spears (forthcoming), although their approach is to reject one adequacy condition outright rather than to neglect it for heuristic reasons.

	First child	Second child
City	20	10
Suburb	25	9

Table 1: Two-child case, from Parfit (2002, p. 83)

I see modelling at work. Here we face the same choice as in section 4, of regarding distributive theories like prioritarianism as mere models, or of characterising them as theories which make contact with particular cases through models of the theory.

In much distributive theorising, the distribution problem is summarised in a table, containing a numerical representation of the distribution problem. Here is a classic case in which Derek Parfit presents a case due to Thomas Nagel.

Nagel imagines that he has two children, one healthy and happy, the other suffering from a painful handicap. He could either move to a city where the second child could receive special treatment, or move to a suburb where the first child would flourish. [...then, quoting Nagel:] I want to suppose that the case has the following feature: the gain to the first child of moving to the suburb is substantially greater than the gain to the second child of moving to the city. [...] To ask my questions, we need only two assumptions. First, some people can be worse off than others, in ways that are morally relevant. Second, these differences can be matters of degree. To describe my imagined cases, I shall use figures. Nagel's choice, for example, can be shown as follows. (Parfit, 2002, pp. 81–83)

Table 1 reproduces his table. There follows this passage, explaining the table.

Such figures misleadingly suggest precision. Even in principle, I believe, there could not be precise differences between how well off different people are. I intend these figures to show only that the choice between these outcomes makes much more difference to Nagel's first child, but that, in both outcomes, the second child would be much worse off. One point about my figures is important. Each extra unit is a roughly equal benefit, however well off the person is who receives it. If someone rises from 99 to 100, this person benefits as much as someone else who rises from 9 to 10. Without this assumption we cannot make sense of some of our questions. We cannot ask, for example, whether some benefit would matter more if it came to someone who was worse off. Parfit (2002, p. 83)

Here we see a clear example of modelling. The vignette contains no numbers, they are introduced as a thinking aid, along with some particular interpretative principles. Importantly, we are told which features to disregard—the precision is an artefact that the modeller, Parfit, wishes us not to impute to the target system. This structure allows

Parfit to create models of the two views, egalitarianism and prioritarianism, in order to investigate their properties in a precise way. For example, egalitarianism, a view about people being equally well off, becomes a view about equality between numbers representing people's welfare. (Once again, we can regard these as models of theories, or models in the absence of full theories.)

To show how all this talk of modelling might help us make progress, I want to consider a more recent debate about prioritarianism. It begins with Otsuka and Voorhoeve (2018), who argue against a form of prioritarianism and in favour of a form of egalitarianism. They use a similar case, introducing some additional structure that it is useful to reflect on. In their case, there is uncertainty about the outcome (in the form of objective, given probabilities). Otsuka and Voorhoeve then make the following qualifications.

We shall assume a measure of utility on which a prospect has higher expected utility for a person just in case it would be preferred for that person's sake after rational and calm deliberation with all pertinent information while attending to her self-interest only. (A person's expected utility is just the probability-weighted sum of her utility in each possible state of the world.) One prospect has the same expected utility as another for a person just in case such deliberation would yield indifference between the two prospects.

[In a footnote to the above:] In other words, we assume that the measure of utility is derived from idealized preferences satisfying the Von Neumann-Morgenstern axioms.[...] More generally, throughout, we assume that orthodox decision theory applies, according to which under risk, a decision-maker ought to maximize the expectation of what he takes to be the relevant value (so that a utilitarian ought to maximize the sum-total of expected utility, a final-utility prioritarian the sum-total of expected priority-weighted utility, etc.). (Otsuka and Voorhoeve, 2018, 9, fn.7)

Here the model is fitted with additional structure, to facilitate yet more precise analysis. The distributive "theories" being discussed (prioritarianism and egalitarianism) do not involve in any essential way these views on utilities, their measurement, and their relation to decision theory. The decision-theoretic link is particularly interesting: decision theory is itself a model; in particular a representational model of agents, which employs various distorting idealisations. Some of these, like the transitivity of preference, are normative assumptions. So, if VNM agents differ from real agents like you and me in this regard, the explanation of that difference is that we ought to be like them. But some of the idealisations are not normative: e.g., these agents have complete preferences. I understand this as a heuristic idealisation: decision theorists know it is not true, but it is included to simplify the analysis by facilitating the use of certain mathematical structures.

Otsuka and Voorhoeve's model also goes beyond the VNM model, by making comparisons between the utilities of individuals possible. Its conclusions are therefore a

complex result of non-normative idealisations about agents, normative idealisations concerning those agents' rationality, additional assumptions to achieve interpersonal comparison of these utilities, and assumptions about how the principles under discussion (prioritarianism and egalitarianism) are realised in the model.

What might it mean to say that a model of this sort has a restricted domain of application? To illustrate this, consider the argument that Otsuka and Voorhoeve make against prioritarianism. They consider a variant of the case with just one child, and uncertainty about whether they will become disabled, and thus how they might benefit from a move to the suburbs or city. Otsuka and Voorhoeve argue that prioritarianism treats risky intrapersonal trade-offs like this as “involving the very same moral calculus as interpersonal trade-offs in which the interests of different people conflict” (Otsuka and Voorhoeve, 2018, p. 9). In so doing, it fails to appreciate the unity of individuals.

A prioritarian might respond that such a case is simply outside of the scope of this model. Prioritarianism, as a model of the good, embodies what an ideally virtuous agent would desire in contexts of *impartiality*, reflecting its primary intended application to questions of distributive social policy. Its domain is multi-person cases, because it is intended to answer the question of how we weigh the competing interests of different individuals.

This is similar to a defence of prioritarianism offered by Adler and Hotug, 2019. But rather than argue for this as a restriction of the domain of their model, they claim that this is simply the domain of ethics itself. “Morality is a framework for resolving interpersonal conflicts; but in a one-person universe there can be no such conflicts” (Adler and Hotug, 2019, p. 121). They do offer a justification of this restriction in scope which seems well-suited to my modeller's answer above: “prioritarians can invoke Otsuka and Voorhoeve's favored explanation of the difference between interpersonal and intrapersonal conflicts, namely an appeal to respectively the separateness and the unity of persons, to motivate such a restriction in scope. According to this line of argument, the unity of persons is decisive in one-person cases...In cases of interpersonal conflict, on the other hand, the separateness of persons comes into play and motivates a prioritarian weighting” (Adler and Hotug, 2019, p. 121). The prioritarian model has a natural domain of interpersonal conflict.

7 Conclusion

Recognising that we are modelling, and adopting modelling more explicitly as a methodology, will bring changes to the practice of moral philosophy. Probably the most significant methodological upshot is a re-evaluation of the role of counterexamples. As a practice, moral philosophy currently thrives on the generation of principles, and their testing against and adjustment in the face of counterexamples, typically in the form of stylised cases where our intuitions contradict the recommendation of the principle. If my characterisation of current work in normative ethics is correct, these counterexamples don't exert the pressure to reject a theory that they are taken to. I don't refer here

to the usual practice of seeking reflective equilibrium, in which a judgement may itself come to be rejected. I mean instead that a counterexample may simply be irrelevant.

If my methodological recommendations are taken up, new ways of working will need to be developed. Rather than being trained to seek counterexamples, graduate students in ethics will need to be trained in modelling: the careful use of idealisation, the analysis of the effects of idealisations through sensitivity analysis, and a new comfort with making progress locally, toward or even in the absence of theory.

References

- Adler, Matthew D and Nils Hotug (2019). “Prioritarianism: A Response to Critics”. In: *Politics, Philosophy & Economics* 18.2, pp. 101–144.
- Anscombe, G. E. M. (Jan. 1958). “Modern Moral Philosophy”. In: *Philosophy* 33.124, pp. 1–19. DOI: 10.1017/S0031819100037943.
- Arras, John (May 18, 2010). “Theory and Bioethics”. In: URL: <https://plato.stanford.edu/archives/win2016/entries/theory-bioethics/#UbiMorPlu> (visited on 04/08/2021).
- Arrhenius, Gustaf (Oct. 2000). “An Impossibility Theorem for Welfarist Axiologies”. In: *Economics and Philosophy* 16.2, pp. 247–266. ISSN: 0266-2671, 1474-0028. DOI: 10.1017/S0266267100000249.
- (forthcoming). *Population Ethics*. Oxford University Press.
- Baier, Annette (1989). “Doing Without Moral Theory”. In: *Anti-Theory in Ethics and Moral Conservatism*. Ed. by Stanley G. Clarke and Evan Simpson. Albany: State University of New York, pp. 29–49.
- Baier, Annette (1985). *Postures of the Mind*. Minneapolis: University of Minnesota Press.
- Bovens, Luc and Stephan Hartmann (2003). *Bayesian Epistemology*. Oxford: Oxford University Press.
- Budolfson, Mark and Dean Spears (forthcoming). “Does the Repugnant Conclusion Have Important Implications for Axiology or for Public Policy?” In: *Oxford Handbook of Population Ethics*. Ed. by Tim Campbell, Krister Bykvist, and Gustaf Arrhenius. Oxford: Oxford University Press.
- Cartwright, Nancy (1983). *How the Laws of Physics Lie*. Oxford University Press. ISBN: 978-0-19-159715-2.
- (1989). *Nature’s Capacities and Their Measurement*. Oxford University Press.
- (1999). *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.
- Chappell, Sophie-Grace (unpublished). “If Not Moral Theory, Then What?” In: *Epiphanies*. manuscript, pp. 1–46.
- Clarke, Stanley G. (1987). “Anti-Theory in Ethics”. In: *American Philosophical Quarterly* 24.3, pp. 237–244.

- Dancy, Jonathan (2017). “Moral Particularism”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2017. Metaphysics Research Lab, Stanford University. URL: <https://plato.stanford.edu/archives/win2017/entries/moral-particularism/> (visited on 09/26/2020).
- Eva, Benjamin and Stephan Hartmann (2019). “On the Origins of Old Evidence”. In: *Australasian Journal of Philosophy* forthcoming in print, pp. 1–14.
- Fotion, Nick (2014). *Theory vs. Anti-Theory in Ethics: A Misconceived Conflict*. Oxford University Press. ISBN: 978-0-19-937354-3.
- 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).
- 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.
- Glymour, Clark N. (1999). “Realism and the Nature of Theories”. In: *Introduction to the Philosophy of Science*. Ed. by Merrilee H. Salmon. Indianapolis and Cambridge: Hackett.
- Hempel, Carl G. (1966). *Philosophy of Natural Science*. Foundations of Philosophy. London: Prentice Hall.
- Leitgeb, Hannes (2013). “Scientific Philosophy, Mathematical Philosophy, and All That”. In: *Metaphilosophy* 44.3, pp. 267–75.
- List, Christian and Laura Valentini (2016). “The Methodology of Political Theory”. In: *Oxford Handbook of Philosophical Methodology*. Oxford: Oxford University Press. URL: <http://personal.lse.ac.uk/list/PDF-files/MethodologyPoliticalTheory.pdf>.
- Louden, Robert B. (July 1, 1990). “Virtue Ethics and Anti-Theory”. In: *Philosophia* 20.1, pp. 93–114. ISSN: 1574-9274. DOI: 10.1007/BF02382586.
- Louden, Robert (1992). *Morality and Moral Theory: A Reappraisal and Reaffirmation*. New York: Oxford University Press.
- McCarthy, David, Kalle Mikkola, and Teruji Thomas (Nov. 3, 2019). *Aggregation for Potentially Infinite Populations without Continuity or Completeness*. arXiv: 1911.00872 [econ]. URL: <http://arxiv.org/abs/1911.00872> (visited on 09/24/2020).
- McKeever, Sean and Michael Ridge (2015). *Obvious Objections*. Oxford University Press. ISBN: 978-0-19-178165-0.
- Morgan, Mary S. and Margaret Morrison, eds. (1999). *Models as Mediators*. Cambridge: Cambridge University Press.
- Nozick, Robert (1981). *Philosophical Explanations*. Oxford: Clarendon Press.
- Nussbaum, Martha (2000). “Why Practice Needs Ethical Theory”. In: *Moral Particularism*. Ed. by Brad Hooker and Margaret Olivia Little. Oxford: Clarendon Press, pp. 234–345.

- Otsuka, Michael and Alex Voorhoeve (2018). “Equality versus Priority”. In: *Oxford Handbook of Distributive Justice*. Oxford: Oxford University Press. URL: [http://personal.lse.ac.uk/OTSUKAM/M%20otsuka%20%20A%20Voorhoeve%20\(final%20w%20author%20edits%202015-06-15%20REVISED%20clean.pdf](http://personal.lse.ac.uk/OTSUKAM/M%20otsuka%20%20A%20Voorhoeve%20(final%20w%20author%20edits%202015-06-15%20REVISED%20clean.pdf).
- Parfit, Derek (1984). *Reasons and Persons*. Oxford: Clarendon Press. ISBN: 0-19-824908-X.
- (2002). “Equality or Priority?” In: *The Ideal of Equality*. Ed. by Matthew Clayton and Andrew Williams. New York: Palgrave Macmillan, pp. 81–125.
- Raphael, D.D. (1974). “The Standard of Morals: The Presidential Address”. In: *Proceedings of the Aristotelian Society* 75, pp. 1–12.
- Reutlinger, Alexander et al. (2019). “Ceteris Paribus Laws”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2019. Metaphysics Research Lab, Stanford University. URL: <https://plato.stanford.edu/archives/win2019/entries/ceteris-paribus/> (visited on 09/26/2020).
- 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. URL: <http://www.jstor.org/stable/20013097>.
- Timmons, Mark (2012). *Moral Theory: An Introduction*. Lanham, MD: Rowman & Littlefield Publishers. ISBN: 978-0-7425-6493-0.
- Titelbaum, Michael G. (forthcoming). “Normative Modelling”. In: *Methods in Analytic Philosophy: A Contemporary Reader*. Ed. by J. Horvath. The PhilPapers Foundation.
- Weisberg, Michael (2007a). “Three Kinds of Idealization”. In: *The Journal of Philosophy* 104.12, pp. 639–659. ISSN: 0022-362X. URL: <http://www.jstor.org/stable/20620065> (visited on 07/10/2018).
- (2007b). “Who Is a Modeler?” In: *The British Journal for the Philosophy of Science* 58.2. URL: <http://www.jstor.org/stable/30115224>.
- Williams, Bernard (1973). “Critique of Utilitarianism”. In: *Utilitarianism: For and Against*. Ed. by Bernard Williams and J. J. C. Smart. Cambridge: Cambridge University Press.
- (1981). *Moral Luck: Philosophical Papers 1973–1980*. Cambridge: Cambridge University Press. ISBN: 978-0-521-24372-8. DOI: 10.1017/CB09781139165860.
- 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.