Synthese (2013) 190:1337–1350
DOI 10.1007/s11229-012-0166-z

Idealisations in normative models

Mark Colyvan

Received: 15 January 2012 / Accepted: 30 July 2012 / Published online: 10 August 2012

Abstract In this paper I discuss the kinds of idealisations invoked in normative theories—logic, epistemology, and decision theory. I argue that very often the so-called norms of rationality are in fact mere idealisations invoked to make life easier. As such, these idealisations are not too different from various idealisations employed in scientific modelling. Examples of the latter include: fluids are incompressible (in fluid mechanics), growth rates are constant (in population ecology), and the gravitational influence of distant bodies can be ignored (in celestial mechanics). Thinking of logic, epistemology, and decision theory as normative models employing various idealisations of these kinds, changes the way we approach the justification of the models in question.

Keywords Formal epistemology · Idealisations · Normativity · Decision theory

1 Introduction

It is very natural to think of various theories in formal epistemology as theories of rationality. Bayesian epistemology can be thought of as a theory of how we ought to organise and update our beliefs; decision theory can be thought of as a theory of how we ought to make decisions; logic can be thought of as a theory of how we ought to draw inferences.[1] Thus construed, the axioms and assumptions of these theories are

[1] Of course, these theories do not have to be thought of as applied in this way. There is nothing wrong, for instance, with thinking of logic as divorced from applications of human reasoning and as a purely mathematical study of the abstract consequence relation. Be that as it may, these three theories draw

M. Colyvan (✉)
Sydney Centre for the Foundations of Science, University of Sydney, A14 Main Quadrangle, Sydney, NSW, Australia
e-mail: mark.colyvan@sydney.edu.au

motivated by canons of rationality and would seem to have little to do with empirical evidence. Indeed, when empirical studies show that human subjects reason in ways that conflict with such theories (Wason 1962; Wason and Johnson-Laird 1972; Kahneman et al. 1982) it is common to see this as evidence for the irrationality, or at least a limitation in the rationality, of humans.

I will argue that although these theories are normative, they are not normative through and through. The main aim of this paper is to begin the task of disentangling the normative elements of these theories from other kinds of idealisations and assumptions. The presence of such non-normative assumptions makes these theories of rationality similar in many respects to models found elsewhere in science. Indeed, I think of the theories in question as *normative models* of belief, decision, and inference. But in order to see the connection with empirical models, I will first need to say a little about scientific models and the idealisations we find in them.

2 Two kinds of model

Scientific models can be either explanatory or descriptive (or some mixture of the two). The former are models designed to tell us how the target system works, without too much attention devoted to getting the details of the system right. The latter are designed to capture specific details of the target system without much attention to why the system is the way it is.[2] Normative models should be contrasted with scientific models; normative models are not supposed to model actual behaviour or explain actual behaviour; rather, they are supposed to model how agents *ought* to act.[3] It will be helpful to have in mind a concrete example of a typical scientific model.

2.1 A scientific model

Consider the Lotka-Volterra model from population ecology. This mathematical model describes the populations of a predator–prey pair via two coupled first-order differential equations:

$$\frac{dV}{dt} = rV - \alpha VP$$
$$\frac{dP}{dt} = \beta VP - qP$$

Footnote 1 continued
at least some of their philosophical significance from the fact that they have natural applications in human reasoning, and it is these applications I am interested in here. Such applications are also why these theories find scientific applications (e.g. Colyvan et al. 2011; Regan et al. 2006).

[2] Think of a city street map as a good example of a descriptive model and evolutionary models as examples of explanatory models.

[3] As I have already mentioned, the prime examples of the normative models are decision theory, Bayesian epistemology, and logic. Indeed, these are usually taken to constitute the "standard picture of rationality" (Stein 1996).

where V is the population of the prey, P is the population of the predator, r is the intrinsic rate of increase in prey population, q is the per capita death rate of the predator population, α is a measure of capture efficiency, and β is a measure of conversion efficiency.

Like many scientific models, the Lotka-Volterra equations both describe the target system and explain certain behaviour of the target system. But for present purposes we can set this complication aside and think of scientific models such as these as descriptive models, describing given target systems. Our real interest is in the idealisations we find in the model and, in particular, why these idealisations are there.

The Lotka-Volterra model is designed to capture the behaviour of the system, but only at a certain level of generality. It cannot faithfully represent all aspects of the target system. It does not, for instance, tell you which of the predators will eat which prey nor how long each given member of a population will live. It is designed to give population-level information about the abundance of the two populations in question. In order to achieve this, the model invokes a number of idealisations. Some of these include:

(a) Population abundance is discrete and yet the model treats it as continuous.
(b) The model treats the growth rates as constants.
(c) The model treats the predator as a specialist (i.e. the predator eats only the prey).
(d) The model treats the prey as having only one predator.
(e) Responses to changes in population abundances are instantaneous.

We find two quite different kinds of idealisations at work here. There are those invoked (purely) for mathematical convenience (e.g. treating population abundance as continuous). I will call these idealisations the *mathematical convenience* kind. These idealisations are usually invoked in order to employ familiar and well-understood mathematical machinery. Although such a move is very common in science, its justification is anything but straightforward. When the mathematics in question is not only convenient but also appropriate in some sense (because, for example, it mirrors the relevant structures in the target system), it is easy to justify. But in other cases, there are clear mismatches between the convenient mathematics and the target system (as in the case under discussion here: with a continuous model of population abundance). In other cases it is not known whether there are structural mismatches between the mathematics and the target system (such as when modelling time with real numbers). I will set such issues aside. Scientists clearly do invoke mathematical machinery for reasons of convenience and that is enough for present purposes.

Now consider another kind of idealisation: those invoked because they are close enough to the way things are. It is hoped that such assumptions result in models that are good enough for the purposes at hand (e.g. growth rates are constant). Call these idealisations the *close enough for jazz* kind.[4] There is an ambiguity here in what it means to be "close enough". Sometimes it might be that the idealisation itself is close to the way things are in the world. For example, some predator might be almost a specialist, in that it rarely eats anything other than the prey represented in the above

[4] Here I am borrowing a phrase from musicians. The suggestion is that the tuning is not perfect but it is close enough for the purpose at hand, namely performing jazz.

model. Alternatively, we might think of an assumption as being close enough in the sense that when included in the model, it delivers results close to what is observed in the target system. Obviously these are not the same; some assumptions might be close in the former sense but due to sensitivity of initial conditions, result in wildly different outcomes. Since it is the closeness to observation that is important in a great deal of scientific modelling, I assume that it is this sense of closeness we have in mind here. Although in particular circumstances, the former sense of closeness can be used as a surrogate for the latter.

Finally, I note that many idealisations are a mixture of both of the above. For example, the idealisation that responses are instantaneous is both mathematically convenient and, for the most part, is close enough to the behaviour of real populations.[5] Be that as it may, it is still useful to distinguish these two quite different motivations for the idealisations in question.

2.2 Normative models

First a note about what I mean by the claim that decision theory, Bayesian epistemology, logic, and the like are normative theories. These theories can, of course, be used to describe or predict what real people will do, and thus construed the theories are descriptive. Indeed, construed this way these theories are just empirical models; they are no different from what we find elsewhere in science (such as the population model in the previous section). Construed as normative theories, however, these same theories are taken to prescribe how we ought to reason, organise our beliefs, and so forth. As I have already mentioned and as is well known, there is a great deal of empirical evidence to suggest that human behaviour is not well described by some of the theories in question. This has led to alternative models which purport to better describe human behaviour. For my purposes, I wish to concentrate on the normative construal of the theories in question: describing ideally rational agents or prescribing our own behaviour.

Of course the two construals just articulated are related. David Lewis (1983, p. 114) suggests that

> [d]ecision theory (at least, if we omit the frills) is not esoteric science, however unfamiliar it may seem to an outsider. Rather, it is a systematic exposition of the consequences of certain well-chosen platitudes about belief, desire, preference, and choice. It is the very core of our common-sense theory of persons, dissected out and elegantly systematized.

[5] It might be argued that the mathematical convenience idealisations are all close-enough-for-jazz or that mathematical convenience idealisations are only successful when they are, in some sense, close enough to how things are. But this is not right, as the idealisation of continuous populations shows. Populations are discrete and properly measured by natural numbers; there is no sense in which real numbers measuring population abundance is close to the truth. The use of real analysis here is purely for mathematical convenience. Of course the resulting model will need to be close enough in the sense of agreeing with empirical results, but the primary motivation here seems to be mathematical convenience: differential equations are, in many way, more convenient than the corresponding difference equations.

So we would expect the descriptive theory to inform the normative theory and since the normative theory is so closely related to the descriptive theory, the normative theory may be set to descriptive purposes. Still, the intuitive distinction between the descriptive construals of these theories and the normative is clear enough, even though that distinction may not stand up to closer scrutiny.

Now consider some of the idealisations or assumptions we find in decision theory (Von Neumann and Morgenstern 1944):

(i) Beliefs come in (continuous) degrees.
(ii) Utilities are dense (and usually represented to be continuous as well).
(iii) Connectedness: there are no incommensurable outcomes.
(iv) Preferences are transitive: if p_1 is preferred to p_2, and p_2 is preferred to p_3, then p_1 is preferred to p_3.
(v) The Archimedian axiom: Whenever an agent has preferences $p_1 < p_2 < p_3$ there will be a lottery (or "mixture") of p_1 and p_3 such that the agent is indifferent between p_2 and the lottery.

There are apparently three different kinds of idealisation here. We find idealisations for mathematical convenience as in the case with scientific models (e.g. that beliefs come in (continuous) degrees). We also find the close-enough-for-jazz kind (e.g. connectedness). But we also find a third kind of idealisation, which I will call *normative constraints*. These are idealisations imposed by rationality; to fail to meet such constraints would apparently be a failure of rationality. An example of such a constraint is (arguably) that preferences are transitive. As before, idealisations may be motivated by more than one of these considerations.

Since it might be initially tempting to think of all of (i–v) as normative, let me say a few words about why this is not the case. Consider the idealisation that degrees of belief are continuous rather than discrete. There could be no normative motivation for this. Indeed, this assumption looks very much like the assumption in population ecology that population abundance is modelled with real numbers, or the assumption in gravitational theory that space-time is modelled with a continuous space-time manifold. Another way of seeing that this is not a requirement of rationality is to consider what failing of rationality we would be witnessing in an agent who had discrete beliefs, modelled by, say, the rational numbers.[6]

Next consider the assumption that all outcomes are connected. This assumptions rules out incommensurable outcomes. Again it can hardly be a requirement of rationality that there are no such outcomes. But this assumption is there because, for the most part, outcomes are commensurable. Connectedness is close enough for many (if not most) purposes. This idealisation also makes the mathematics more tractable, but I take it that its primary motivation is of the close-enough-for-jazz kind. If, however, you are committed to the view that there is widespread incommensurability, you might look upon this idealisation as more a case of mathematical convenience. For present purposes, it does not matter; what matters is that connectedness is not a requirement of

6 Another example of this kind of idealisation is Ramsey (1928) assumption of a morally neutral act in his presentation of decision theory. Such an act is postulated purely for mathematical convenience and bears similarity to point masses in celestial mechanics. Thanks to Jim Joyce for this suggestion.

rationality. It is more like the kind of idealisation one finds in scientific models about, for example, the topology of space-time.

I am inclined to think that of the idealisations above, only (iv)—the transitivity of preferences—is defensible as a normative constraint. This defence is via a money-pump argument (Ramsey 1928).[7] The idea here is that if an agent violates transitivity you can financially exploit this violation. You can arrange things so that the agent in question effectively gives you money with nothing in return, and this, in turn, is taken to be a hallmark of irrationality. All the other idealisations are arguably, there for reasons of mathematical convenience or because they are close enough for jazz, or both.

Now consider some of the idealisations or assumptions we find in classical logic:

(1) There are only two truth values.
(2) The valuation relation is a *function*.
(3) Every proposition in the language is assigned a truth value.
(4) The logical particles are truth functional.

Here we find the same three kinds of idealisation. We have idealisations for mathematical convenience, such as that the the connectives are truth functional. There are the close-enough-for-jazz idealisations such as that every proposition is assigned a truth value. Then there are normative constraints such as the behaviour of at least some of the connectives (such as the truth table for $\wedge$). And as before, many of the idealisations might be motivated by a mixture of more than one of the three considerations.

The existence of non-normative idealisations in logic is clear but is not often stressed. Take the truth functionality of the connectives, for example. There is nothing about rationality that demands this. It is clearly assumed for reasons of tractability (or mathematical convenience). After all, why would we expect that the conditional should be truth functional? Assuming truth functionality of all connectives just makes life easier. Indeed, without this idealisation it is initially hard to see how to get started with formal logic. Of course, we now have non-truth-functional connectives (such as Stalnaker conditionals, modal operators, and intuitionistic negation) but there is no denying the simplicity of purely truth-functional connectives. It is this simplicity that is a prime motivation for the idealisation of truth functional connectives in classical logic.

The case for the truth table for classical conjunction is rather different. A case can be made that this truth table forces itself upon us—it constitutes the meaning of the English "and" (at least in the context of a two-valued logic). Moreover, anyone who violates the classical truth table by, say, believing "A and B" but not believing B, is either irrational or is not grasping the meaning of "and".[8]

Similarly, the idealisation that every proposition in the language takes a truth value cannot be a requirement of rationality. After all, future contingents, vagueness,

[7] Interesting arguments against the rationality of the transitivity of preferences have also been advanced (Armstrong 1939; Luce 1956) but still there is a case to be made for transitivity as a requirement of rationality. I set aside the plausibility of the respective cases.

[8] There are some interesting alleged counterexamples to the commutativity of conjunction. For example, "she had a sleep and went to work" differs in meaning from "she went to work and had a sleep". Such examples may be seen as casting doubt on the commutativity of conjunction as a normative constraint.

fictional discourse and the like all present prima facie examples of propositions which fail to take truth values. In any case it cannot be a failure of rationality to hold that future contingents, say, lack truth values. Nevertheless, the idealisation that all propositions take truth values is a good one—it holds much of the time and this is why it is invoked. For many purposes, it is perfectly innocent. In this regard it is very similar to the connectedness assumption in decision theory. Depending on how wide spread you think such truth-value gaps are, the idealisation in question can be thought of as either close enough for jazz or one of mathematical convenience. Either way, it is not dictated by rationality.[9]

3 Questioning assumptions of models

So far I have argued that there are three quite different motivations for idealisations and assumptions in normative theories and not all these motivations derive from rationality considerations—some are there for reasons of mathematical convenience or because they are close enough for a given purpose. The different possible motivations turns out to be of significance when we wish to criticise a model. It will not do to criticise a scientific model as inadequate simply because it invokes idealisations or false assumptions. Consider, for example, a criticism of a population model that proceeded by pointing out that populations are discrete, not continuous. This gives us no reason to reject the model. The idealisation is indeed false but it is there, not because it is thought to be true, but because of the mathematical convenience it affords. Other criticisms of scientific models, however, can be made to stick. For example, we can question the "close enough for jazz" idealisation in situations where it is not jazz: for example, a population ecology model assuming a constant population growth rate when in fact the growth rate is changing dramatically.[10] It is important to keep track of the status of an idealisation before launching an attack on it or on the model it belongs to. This much is well known. Now let us turn to questioning idealisations in normative models.

Just as in descriptive models, we can criticise a close-enough-for-jazz assumption in a normative model when it is not jazz. For example, when dealing with fictional discourse, it might be argued that the assumption that every proposition takes a truth value leads to absurd results and thus should be abandoned. Even idealisations introduced for mathematical convenience can be criticised: they can skew the way we look at the phenomenon in question. For instance, the idealisation that beliefs can be adequately represented by a real-valued function biases the debate away from more

[9] Another important class of normative theories not discussed here are ethical theories. My main reason for setting ethical theories aside is that they are usually not represented formally and my target here is squarely on formal theories. Attempts to formalise ethical theories in order to combine them with formal decision theory (Colyvan et al. 2010; Louise 2004; Oddie and Milne 1991) might provide an opportunity to investigate formally-represented ethical theories along similar lines.

[10] If you prefer, you might think of cases like this as cases where the idealisation is not close enough—not even for jazz. It does not much matter how the metaphor is cashed out here; what is important is that the idealisation in question, is benign in some settings (when the growth rate is not changing too much) and in other settings it is not.

complicated belief functions [such as those suggested by Shafer (1976) and Walley (1991)]. And the idealisation that the valuation relation of logic should be a function biases the debate away from various non-classical logics that allow propositions to be associated with more than one truth value (Priest 2008). And of course, we can also criticise normative idealisations (e.g. transitive preferences and the truth table for $\wedge$) by demonstrating that an agent can reasonably violate the idealisation in question.[11]

Sometimes the problem is that unintuitive consequences follow from the combined idealisations of the model. In such cases it is hard to know where to place the blame. What idealisations, for instance, are responsible for, the Pasadena paradox (Nover and Hájek 2004) or the Newcomb problems of decision theory (Nozick 1969)? What are the idealisations responsible for the sorites paradox or for explosion in classical logic?[12] In such cases, does the fault lie in some of the apparently harmless mathematical convenience idealisations skewing the picture or in some of the normative or "close enough for jazz" assumptions? It can be very difficult to isolate the root of the problem in such cases. The blame can be directed at a number of different assumptions.[13]

4 Case studies

It is worthwhile looking at a couple of case studies to get a feel for how the debates proceed in these more complicated cases. I will consider one case study from rational belief theory and one from philosophy of logic. The point is to see how thinking in terms of the status of the various idealisations can help us to adjudicate the debates and we see how even constraints of rationality can be questioned.

4.1 Cox's theorem

Cox's theorem (Cox 1946, 1961) is a mathematical result used to defend a Bayesian interpretation of the probability calculus. It tells us that any measure of belief is isomorphic to a probability measure. It rests on a number of explicit and implicit assumptions. The explicit assumptions are: (i) an agent's belief in $\neg P$ is a function of her belief in P and (ii) an agent's belief in $P \wedge Q$ is a function of her belief in P given Q and her belief in Q. The implicit assumptions include: (iii) belief is a real-valued function, (iv) the underlying logic is classical propositional calculus, (v) the following substitution principle holds: if an agent believes P to some degree, b, and

[11] Of course there will always be other models that relax the contentious idealisations. In science we build more complicated models and in the normative domain we have alternative logics, decision theories, and epistemologies. But these alternative models still invoke idealisations—we just have different idealisations.

[12] Explosion is an unintuitive feature of classical (and other) logic(s) that everything follows from a contradiction. The sorties paradox will be discussed in more detail in one of the case studies to follow.

[13] This also happens in scientific models. It is not always easy to see what the faulty assumption is in a failed model. Even apparently innocent mathematical convenience idealisations can contribute to the failure of scientific models. See Bueno and Colyvan (2011) for more on the relationship between mathematical models and physical systems.

P is logically equivalent to Q, then she believes Q to degree b and (vi) the function in (ii) is twice differentiable.

What we have here is a model of the structure of rational belief. We then derive results in this model and thereby take to have shown that these results are true of rational belief. Cox's theorem is a representation theorem. It shows that if belief has the assumed structure, it can be modelled by the probability calculus. This is one of many similar theorems that lend support to a subjective interpretation of the probability calculus. But some take matters further and conclude that the theorem lends support to the thesis that the *only* coherent treatment of uncertainty is probabilistic (e.g. Lindley 1982; Jaynes 1988; Van Horn 2003). But putting the theorem to work towards such ends suggests that the assumptions (both explicit and implicit) are all demanded by rationality. This, however, is problematic.

I have discussed the problems with such applications of Cox's theorem elsewhere (Colyvan 2004, 2008). Here I will mention just a couple of issues that are relevant for present purposes. Consider assumption (iv). The assumption that the underlying logic is classical is not justified if Cox's theorem is to be used to argue against non-classical systems of belief representation. And the assumption that the function in (ii) is twice differentiable seems like an innocent mathematical idealisation, but it rules against max and min functions used in some infinite-valued logics. Moreover, excluded middle and double negation elimination can be questioned in contexts where there is uncertainty due to vagueness or fictional entities. This is not to say that one can never invoke such idealisations. Indeed, such idealisations seem perfectly reasonable when Cox's theorem is used as a way of justifying a subjective interpretation of the probability calculus in non-fictional domains, where there are no concerns about vagueness. The innocence or otherwise of the assumptions depends on what the theorem is being used for. Just as in empirical models, the idealisations of the model depend on the purpose of the model.

4.2 The sorites paradox

Recall that the sorites paradox invokes vague predicates to raise problems for classical logic. In its canonical form the paradox arises from the following two premises: "an m-grained collection (for some suitably-large m) of sand is a heap", and "if an n-grained collection is a heap, so too is a $(n-1)$-grained collection. The second premise is true because of the vagueness of the predicate "is a heap". By repeated applications of the valid argument form modus ponens, we can draw the apparently false conclusion that a one-grained collection of sand is a heap. By a symmetric argument, we can prove that a m-grained collection of sand is not a heap. The sorites argument thus delivers genuinely contradictory conclusions. Of course, there is nothing special about the predicate "is a heap". We can construct a sorites paradox with any vague predicate.

The paradox is generated by the combination of classical logic and vague predicates. The debate about how to resolve the problem is ongoing, with many alternative proposals defended. One obvious proposal is to blame the assumption of classical logic that there are only two truth values.[14] I suggested previously that this idealisation is motivated by mathematical convenience. Once this idealisation is dropped, the question of the number of truth values is opened up. Answers range from three truth values in K_3 to infinitely-many truth values in the Łukasiewicz system $Ł_\infty$.

The supervaluationalists approach (Fine 1975) rejects the conditional premise, but denies that there is a particular cut off between heaps and non-heaps. In order to do this a non-classical logic is invoked—one that preserves the classical principle of excluded middle by accepting odd behaviour from the disjunction connective.[15] This looks like the rejection of a normative constraint, namely, the usual truth table for disjunction. Also sacrificed is the mathematical convenience of truth functionality for the logical connectives. The case for the supervaluational approach usually proceeds by stressing the senses in which supervaluational logic is nearly classical[16] and motivating the departures from classical logic as being minimal and forced by the paradoxical argument. It rarely proceeds by directly challenging the truth table for disjunction. The rejection of the truth table for disjunction is seen as the price to be paid for an otherwise elegant solution to the paradox.

Other proposed solutions to the sorites paradox reject the validity of modus ponens, as is the case with subvaluationalism (Hyde 1997).[17] This again involves the rejection of an apparently normative constraint. By contrast, the epistemicist solution (Sorensen 1988; Williamson 1994) holds firm with classical logic. Instead, the epistemicist accepts the counterintuitive claim that there is a cut off between heaps and non-heaps, but that the location of this cut off is unknowable.

Everyone in the debate is forced to give up something. Most give up classical logic, sometimes by relaxing mathematical idealisations, such as bivalence. This move is similar to the way scientific modellers relax idealisations of mathematical convenience when these idealisations cease to be convenient or start delivering poor results. Others in the sorties debate are prepared to give up what are apparently normative idealisations such as a detachable conditional.[18] The case for such a move is usually presented in terms of overall elegance of the theory (or at least in terms of parity with other elegant solutions (Hyde 1997)), rather than by direct attacks on any of the normative idealisations themselves.

It might seem that giving up idealisations in place for mathematical convenience should be more readily acceptable than giving up those with normative motivation. One would thus expect that multi-valued logic would be easier to accept than, say, supervaluationalism. But in philosophical circles, at least, this seems not to be the case. Both multi-valued logics and supervaluationalism remain popular and are thought to be tenable. The normative idealisations although perhaps a little more firmly entrenched are not shielded from revision by virtue of their normative status.

[14] For example, see Smith (2008) for a recent, detailed philosophical defence of such a solution.

[15] Recall that in supervaluational logic, a disjunction can be true without either disjunct being true, and an existentially-quantified formula can be true without any true instances.

[16] It preserves all the theorems of classical logic, for instance.

[17] Subvaluationalism is the dual of the familiar supervaluationalism but identifies truth with *true in at least one precisification* (rather than in all precisifications).

[18] A detachable conditional is one that supports modus ponens.

5 The normative and the empirical

I have argued that normative models, such as rational-belief models and rational decision models, are not so different from models found in empirical science. The former have some genuinely normative idealisations, whereas the latter do not. But not all the idealisations in normative models are normative; there are also the familiar mathematical-convenience and close-enough-for-jazz idealisations, as in the models found in empirical science. When a model fails to deliver what is expected of it, this can be for a number of reasons. The falsity of one or more of the assumptions is never the full story, and sometimes is beside the point. After all, false assumptions are found in all models. The question is why some, but not all, models perform the tasks we expect of them. This is an important question and one without a satisfactory answer in the scientific-modelling literature, but an extra degree of difficulty is added when we consider normative models.

This naturally leads to the question of what it means for a normative model to be adequate. In the case of scientific models it is clear enough: a model is adequate just in case it delivers the desired explanatory power, predictive accuracy, or whatever the model was designed for. Moreover, the success or otherwise of models for many such purposes will be empirically testable. But what is it for a normative model to be adequate? As we have learned from the case studies, we need to pay attention to the intended purpose of the model and the level of detail it is supposed to provide. This is the same for empirical models. The primary purposes of normative models seems to be to deliver good advice about decisions, inferences, the structure of beliefs and the like—at some appropriate level of abstraction. But if this is the purpose of normative models, it is hard to see how they are empirically testable. To test the advice given by a normative model, we would already need to know what constitutes good advice in the circumstances in question. Alternatively we might take the model itself as delivering good advice simply because the advice is delivered by the model in question. We find ourselves with a kind of Euthyphro dilemma. In either case, it is hard to see what would constitute an empirical test of the model.

There is a way of understanding normative models that has emerged from at least the second of the two case studies and which gives us the resources to tackle the question of empirically testing normative models. We treat these models holistically. That is, rather than defending each idealisation and constraint in the model, we judge the model as a package. Here we follow the lead of modern set theory, where the axioms of, say, ZFC are justified by a “top-down” approach—by the fruits of the theory as a whole (Russell 1907; Gödel 1947). In set theory, the fruits include, delivering elegant and instructive proofs of various theorems, opening up new areas of inquiry, and the like. In case of the normative theories, the fruits might be thought to include: meshing with intuitions about what to do in given situations, shedding light on other situations where intuitions fail, and perhaps facilitating elegant representation theorems (where appropriate).

It is interesting to note that the holistic approach is taken in set theory because of the failure of the “bottom up” approach of justifying each axiom by a priori means. It was this latter approach that led to the acceptance of an inconsistent theory: naive set theory with its intuitively plausible but contradictory unrestricted comprehension axiom (Giaquinto 2002). Of course the theories we are considering here are different in various respects, so this is merely an analogy.[19] The point is that there is a way of justifying an a priori theory such as set theory that does not depend on an axiom-by-axiom justification. Similarly, holistic justifications of decision theory, logic and formal epistemology are live alternatives.[20]

Such a "top-down" approach to justifying normative models means that they are not fully normative. They are not fully normative in the sense that not every idealisation has an a priori normative justification. This understanding of normative models provides yet another way in which they are similar to empirical models: as with empirical models we are interested in whether a given model—as a whole—is adequate for the purpose for which it was designed. There is less interest in justifying the various false assumptions (except, perhaps, to defend some of the "close-enough-for-jazz assumptions as indeed being close enough). So too with normative models, according to the proposal being advanced here. Quine (1980, p. 41) argued that "our statements about the external world face the tribunal of sense experience not individually but only as a corporate body". The current suggestion is that our normative models also function as corporate bodies.

Are normative models empirically testable on this view? In rejecting a priori normative justifications for every facet of the model, I am at least opening up the possibility for empirical testing (and for a role for experimental philosophy). The details of how such testing would proceed, however, will be hard to spell out. After all, the failure of even a large proportion of the population to reject, say, the logical fallacy of affirming the consequent says nothing about the adequacy of classical logic. The more reasonable conclusion from any such study would seem to be that people are not good at deductive reasoning. But empirical testing is not completely irrelevant. If no-one, including expert logicians, agrees with the logical model, one would be entitled to question whether the model was delivering the right results. Where we draw the line between empirical results telling against the model or against the rationality of the test population is a subtle matter. The adjudication in question needs further work, but it is clear that not all conflicts between empirical results of actual human behaviour and predictions of the normative theory should be interpreted as human irrationality.[21] Sometimes the normative model should take the blame. And here even the normative idealisations of the model are not immune from revision.

Acknowledgments The author gratefully acknowledge the assistance of Rachael Briggs, Andy Egan, Alan Hájek, Jim Joyce, Steven Robertson, Katie Steele, and three anonymous referees for this journal. The author is also indebted to audiences at the University of Michigan, the University of Otago, the Australian National University, and Lund University, where he received many constructive comments on earlier versions of this paper. The author is especially grateful to the organisers and fellow participants in the conference "Formal Epistemology Meets Experimental Philosophy" at Tilburg University in September

[19] Although the case with logic is very similar: an intuitive notion of truth in classical logic led to contradiction in the form of the liar paradox.

[20] Further case studies of particular debates in these areas will help determine how widespread such holistic approaches are and help bollster the defence of such methods.

[21] This is important when interpreting the results of empirical work by, for example, Kahneman et al. (1982).

2011. Work on this paper was jointly supported by an Australian Research Council Discovery Grant (Grant Number: DP0879681) and project funding from the Australian Centre of Excellence for Risk Analysis.

References

Armstrong, W. E. (1939). The Determinateness of the utility function. *Economic Journal, 49*, 453–467.

Bueno, O., & Colyvan, M. (2011). An inferential conception of the application of mathematics. *Noûs, 45*(2), 345–374.

Colyvan, M. (2004). The philosophical significance of Cox's theorem. *International Journal of Approximate Reasoning, 37*(1), 71–85.

Colyvan, M. (2008). Is probability the only coherent approach to uncertainty?. *Risk Analysis, 28*(3), 645–652.

Colyvan, M., Cox, D., & Steele, K. (2010). Modelling the moral dimension of decisions. *Noûs, 44*(3), 503–529.

Colyvan, M., Justus, J., & Regan, H. M. (2011). The conservation game. *Biological Conservation, 144*(4), 1246–1253.

Cox, R. T. (1946). Probability frequency and reasonable expectation. *American Journal of Physics, 14*, 1–13.

Cox, R. T. (1961). *The algebra of probable inference*. Baltimore: Johns Hopkins Press.

Fine, K. (1975). Vagueness, truth, and logic. *Synthese, 30*, 265–300.

Giaquinto, M. (2002). *The search for certainty: A philosophical account of foundations of mathematics*. Oxford: Clarendon Press.

Gödel, K. (1947). What is Cantor's continuum problem? (reprinted, revised and expanded). In P. Benacerraf & H. Putnam (Eds.), *Philosophy of mathematics selected readings* (2nd edn, pp. 470–485). Cambridge: Cambridge University Press.

Hyde, D. (1997). From heaps and gaps to heaps of gluts. *Mind, 106*, 641–660.

Jaynes, E. T. (1988). How does the brain do plausible reasoning?. In G. J. Erickson & C. R. Smith (Eds.), *Maximum entropy and bayesian methods in science and engineering*. Dordrecht: Kluwer.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Lewis, D. (1983). *Philosophical papers* (Vol. 1). Oxford: Oxford University Press.

Lindley, D. V. (1982). Scoring rules and the inevitability of probability. *International Statistical Review, 50*, 1–26.

Louise, J. (2004). Relativity of value and the consequentialist umbrella. *The Philosophical Quarterly, 54*, 518–536.

Luce, R. D. (1956). Semiorders and a theory of utility discrimination. *Econometrica, 24*, 178–191.

Nover, H., & Hájek, A. (2004). Vexing expectations. *Mind, 113*(450), 237–249.

Nozick, R. (1969). Newcomb's problem and two principles of choice. In N. Rescher (Ed.), *Essays in honour of Carl G*. Dordrecht: Hempel. Reidel.

Oddie, G., & Milne, P. (1991). Act and value: Expectation and the representability of moral theories. *Theoria, 57*, 42–76.

Priest, G. (2008). *An introduction to non-classical logic: From if to is, 2nd ed*. Cambridge: Cambridge University Press.

Quine, W. V. (1980). *From a logical point of view, 2nd ed.* (pp. 20–46). Cambridge: Harvard University Press.

Ramsey, F. P. (1928). Truth and probability. In R. B. Braithwaite (Ed.), *The foundations of mathematics and other logical essays*. London: Routledge and Kegan Paul.

Regan, H. M., Colyvan, M., & Markovchick-Nicholls, L. (2006). A formal model for consensus and negotiation in environmental management. *Journal of Environmental Management, 80*(2), 167–176.

Russell, B. (1907). The regressive method of discovering the premises of mathematics (reprinted). In D. Lackey (Ed.), *Essays in analysis* (pp. 272–283). London: George Allen and Unwin.

Shafer, G. (1976). *A mathematical theory of evidence*. Princeton: Princeton University Press.

Smith, N. J. J. (2008). *Vagueness and degrees of truth*. Oxford: Oxford University Press.

Sorensen, R. (1988). *Blindspots*. Oxford: Clarendon Press.

Stein, E. *Without good reason: The rationality debate in philosophy and cognitive science*. Oxford: Clarendon Press.

Van Horn, K. S. (2003). Constructing a logic of plausible inference: A guide to Cox's theorem. *International Journal of Approximate Reasoning, 34*, 3–24.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.

Walley, P. (1991). *Statistical reasoning with imprecise probabilities*. London: Chapman and Hall.

Wason, P. C. (1962). *Psychological aspects of negation: An experimental enquiry and some practical applications*. London: Communication Research Centre, University College London.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. Cambridge: Harvard University Press.

Williamson, T. (1994). *Vagueness*. London: Routledge.