

Normative Models and Their Success

Philosophy of the Social Sciences
2021, Vol. 51(2) 123–150
© The Author(s) 2020



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0048393120970908
journals.sagepub.com/home/pos



Lukas Beck¹ and Marcel Jahn²

Abstract

In this paper, we explore an under-investigated question concerning the class of formal models that aim at providing normative guidance. We call such models *normative models*. In particular, we examine the question of how normative models can *successfully* exert normative guidance. First, we highlight the absence of a discussion of this question—which is surprising given the extensive debate about the success conditions of descriptive models—and motivate its importance. Second, we introduce and discuss two potential accounts of the success conditions of normative models. Our tentative conclusion is that the second account is more promising.

Keywords

models, normativity, idealizations, decision theory

1. Introduction

There are two fundamental roles that formal models can play: a descriptive and a normative one (Buchak 2013; Thoma 2019). As descriptive devices, they aim at representing a particular part of the world in order to facilitate explanations and predictions. Scrutinizing models in this descriptive sense has been subject of a longstanding philosophical debate (Elgin 1983; Frigg

Received 11 October 2020

¹University of Cambridge, UK

²Humboldt-Universität zu Berlin, Germany

Corresponding Author:

Lukas Beck, Department of History and Philosophy of Science, University of Cambridge,
Free School Lane, Cambridge CB2 3RH, UK.

Email: lb760@cam.ac.uk

and Nguyen 2016; Hesse 1963; Khosrowi 2018; Mäki 2009; Morgan and Morrison 1999; Weisberg 2013). The aim of this debate is to reconstruct why successful descriptive models are successful by illuminating the relationship in virtue of which the model achieves an adequate representation of its target. As normative devices, on the other hand, models aim at providing normative guidance to agents. If a model is taken to play this kind of role, we call it a *normative model*. They are particularly salient in decision theory, economics, and formal epistemology. Surprisingly, normative models as a general phenomenon are severely under-investigated as of now. Given the vital role of normative models in policy-making and applied ethics, it is, therefore, quite puzzling why there is little systematic discussion of normative models.¹

In this paper, we aim at remedying this situation by addressing what we take to be the core issue of normative modeling: how do normative models meet their aim of *successfully* exerting normative guidance toward agents? More specifically, we wish to engage the question in virtue of which property such models can achieve this aim. This property is what we refer to as the *normative source* of the model. As we will illustrate in the paper, pinning down the normative source of a normative model is not trivial. This is mainly due to the fact that normative models involve what, following Colyvan (2013), we call *descriptive idealizations*, that is, idealizations that make false factual claims about the agents the model is targeting. Thus, the puzzling question arises how models involving such false descriptions of agents can provide normative guidance to them.

In order to make progress toward addressing this puzzle, we lay out and discuss two different accounts of the normative source of a normative model. The first account relies on the intuitive idea that normative models are not too different from another device that affords us with normative guidance: normative arguments. In particular, this account assumes that normative models and normative arguments are analogous concerning their normative source: while the normative source of a normative argument is grounded in its premises, the normative source of a normative model is grounded in its assumptions. We highlight several challenges for this account that arise due to the fact that normative models involve descriptive idealizations in their assumptions.

These challenges, while not conclusively rejecting the first account, motivate us to explore an alternative. According to the second, alternative account,

¹Philosophers have recently started to ask how scientific tools are apt to provide us with normative guidance (see, for instance, Alexandrova 2018; Hersch 2020). Yet, this debate has, so far, not focused on normative models.

the ability of a normative model to successfully exert normative guidance consists in its function to extend normative justification from cases in which we have firm normative verdicts to cases in which we lack those. Roughly speaking, under this account, the normative source of a normative model consists in its ability to capture patterns in our already justified normative verdicts and to, subsequently, project them onto cases in which we lack justified normative verdicts. Thus, under the second account, the normative source of normative models crucially depends on normative verdicts whose justification has to be provided independently of the model. We highlight that this alternative account, while only granting normative models a relatively weak form of normative guidance, is able to evade the challenges raised against the first account. It, therefore, provides a fruitful starting point for further inquiries. The overall aim of our paper is to pursue some first steps in exploring the success conditions of normative models as a so far relatively under-investigated topic.

The paper proceeds as follows. Section 2 outlines the general phenomenon of normative models. Section 3 highlights that the current literature has little to offer when it comes to illuminating the success conditions of normative models. Section 4 and 5 discuss the two accounts of the normative source of normative models.

2. Normative Models

2.1. Basics

We wish to start by outlining how we conceive of normative models and explicate why we hold that there is a puzzle concerning their guidance function. What is peculiar about models in general is their use of idealizations, which are deliberate distortions or simplifications of their target-system (confer Weisberg 2013; Potochnik 2017). Most models are devices that aim at representing a certain part of the world, their so-called target-system, in order to facilitate descriptive aims such as explanations or predictions. If a model is primarily employed to achieve such descriptive aims, we call it a *descriptive model*. Examples of such models are the Lotka-Volterra model of predator-prey interaction in biology, the IS-LM model in macroeconomics, and the ideal-gas model in statistical mechanics.

Yet, there is also another class of models that—although they may, on some occasions, be employed for descriptive purposes—are, on other occasions, taken to provide us with normative guidance (see also Buchak 2013; Thoma 2019). Indeed, it is often the case that one and the same model is, without any change to its formal structure, employed in both projects. For

instance, a decision-theoretic model can both aim at representing an agent (i.e., how the agent makes decisions) but also at recommending courses of action (i.e., how the agent should act). In our terminology, a model counts as a *normative model* if it is employed in the second way, that is, if it is taken to provide normative guidance.

More specifically, a normative model provides normative guidance if it issues normative verdicts, that is, recommendation for how to act, and an agent can appeal to such verdicts in order to determine her course of action.² Moreover, as we understand it, a normative model *successfully* exerts normative guidance if the agent's appeal to the model's verdicts would be justified in the following sense: by following the model's verdicts, the agent would act in accordance with a normative reason provided by the model.³ Importantly, the relevant notion of a normative reason pertains to the class of reasons that determine what the agent targeted by the model ought to do.⁴ To put it differently, the model needs to be such that it can bestow some course of action with a normative pull.

As we will see, it is unclear how normative models manage to *successfully* exert normative guidance. Roughly speaking, this is due to the fact that

²Note that descriptive models can, of course, also be used for normative purposes without qualifying as a normative model in our sense. For instance, given that an agent already knows where she *should* go, she could rely on a map she believes to be accurate in order to find her way to her destination. In such a case, however, the model seems to only facilitate her correct means-end reasoning that is sometimes relevant for normative purposes, but the model does not itself issue a normative verdict about what she should do. We thank an anonymous reviewer for the mentioned example that hopefully helped us clarify this issue.

³Three clarificatory remarks. First, due to other normative considerations besides the reasons provided by the model, the agent may overall have more reason to act against the verdicts of the model. Second, note that our notion of normative guidance is weak as it does not assume that the agent's psychological motivation for performing an action is captured by the structure of the model. Yet, it is still a form of guidance insofar as the agent can appeal to the model's verdicts in order to determine courses of action that would be justified by *pro-tanto* normative reasons. Third, we are aware of the fact that our notion of normative guidance differs from how it is sometimes used in the meta-normative literature where authors assume that an agent is normatively guided by, say, a normative theory if it captures something crucial about the agent's psychological make-up during, or prior to, action (for a recent discussion, see Fox 2019).

⁴To see that not all reasons have such a kind of normative force, just consider that there could be a reason for an agent to, let us say, have continuous preferences because it is merely "nice" to have such preferences. Yet, such an *enticing reason* (see Dancy 2004) may not bear on whether the agent ought to have continuous preferences.

normative models represent the agents they are supposed to guide—that is real-world agents—in a highly idealized way. This prompts the puzzling question of how such false representations can be of any normative significance to real-world agents. However, to make the problem explicit, we have to say more about the kinds of idealizations that normative models employ.

2.2. Two Types of Idealizations

There are two kinds of idealizations introduced by normative models (see also Colyvan 2013). On the one hand, there are what we call *descriptive idealizations*. These are false factual statements about the model's target and are usually introduced with the objective of making the model more tractable by reducing complexity or to afford mathematical convenience. Importantly, the primary justification for invoking descriptive idealizations has to do with the specific aims of the model at hand. For instance, in the case of a predictive model, a particular idealization could be invoked because it makes the model more tractable and, thereby, allows us to derive predictions.

This kind of idealization has to be distinguished from what we call *normative idealizations*. A normative idealization picks out a general fact about how an agent *should* behave (or how her attitudes *should* look like). Thus, normative idealizations are essentially those which the agent has reason to live up to anyway, that is, independent of the idealizations being part of the model. Note that, according to this definition, descriptive models can involve normative idealizations as well. However, in the case of descriptive models, the distinction between the two types of idealizations becomes irrelevant as descriptive models do not serve a normative purpose. Put differently, it is irrelevant to their purposes whether their assumptions are normatively justified.⁵

Given these considerations, we hold that normative idealizations are unproblematic with respect to elucidating the normative guidance function of normative models. After all, normative idealizations are closely tied to an

⁵One may find the term “normative idealizations” unintuitive because models' idealizations are commonly understood as being false but, one might think, normative idealizations describe *true* normative facts. However, we hold that our use of normative idealization is not at odds with how idealizations are commonly understood because normative idealizations are, by and large, false descriptions in the sense that they aim at representing how things ought to be and not how they actually are. This does, of course, not preclude that some agents already live up to the normative propositions stated in the normative idealizations. We thank an anonymous reviewer for pushing us to clarify this issue.

agent's normative reasons.⁶ However, in contrast to normative idealizations, we take descriptive idealizations to generate a puzzle concerning the guidance function of normative models: since descriptive idealizations simply lead to false representations of their targets—that is, the real-world agents the model is supposed to guide—the question arises of how models involving such falsehoods could generate normative verdicts the agent has reason to follow.

Before tackling this question head-on, we shall first take a closer look at particular normative models in order to illustrate that they, in fact, involve a mixture of descriptive and normative idealizations. Thereafter, we will say more about the sort of normative guidance provided by normative models. This will put us in a better position to address the question of how normative models can *successfully* exert normative guidance. We will mainly focus on models based on *expected utility theory* (for short, EUT).

EUT models put forth the following assumptions:

- (i) *Completeness*: For every option A and B, either A is (weakly) preferred to B or B is (weakly) preferred to A, or both.
- (ii) *Transitivity*: If A is preferred to B, and B is preferred to C, then A is preferred to C.
- (iii) *Independence of Irrelevant Alternatives*: Whenever an agent prefers A to B, then, for every probability $0 \leq p \leq 1$ and option C, she will also prefer $p \cdot A + (1-p) \cdot C$ to $p \cdot B + (1-p) \cdot C$.
- (iv) *Continuity*: Whenever an agent prefers A to B to C, there is a probability p such that the agent is indifferent between $p \cdot A + (1-p) \cdot C$ and B.

It has been frequently argued that agents' actual preferences fail to satisfy each of these assumptions (see, for instance, Bruni and Sugden 2007; Tversky 1969; Tversky and Kahneman 1979). Hence, it is safe to say that all of them should be viewed as idealizations. However, out of these assumptions, only Transitivity can confidently claim to be a normative idealization. It is frequently advocated as a requirement of rationality (Broome 2013), and one may appeal to money-pump arguments to establish that one should have transitive preferences. Consequently, it is plausible to assume that one *should* have transitive preferences.

⁶Note that we are not committed to a precise conception of normative reasons. So, we take no stance as to whether reasons are, say, primitive entities (Scanlon 2014) or whether reasons should be explicated by reference to some other concept as suggested by, for instance, Humeanism (Schroeder 2007) or value-based theories (Maguire 2016).

Whether the Independence of Irrelevant Alternatives assumption should count as a requirement of rationality has been hotly debated ever since it was put forward by Von Neumann and Morgenstern (1944). However, independent of its normative status, it enables us to represent agents' preferences via a cardinal as opposed to an ordinal utility function. Therefore, it allows us to perform certain mathematical operations like comparing differences between utilities that would not be possible in its absence. Hence, even if it is not supported by a normative reason, there are pragmatic rationales for introducing this idealization (like providing us with mathematical convenience).

Irrespective of the precise status of the Independence assumption, the other two assumptions—Completeness and Continuity—usually do not come with any model-independent normative reasons that would render them normative idealizations. First, the Completeness assumption basically tells us that agents do not judge outcomes as incommensurable or on par (Chang 1997). While this may be true for some contexts, in other contexts in which EUT models are employed it will misrepresent the evaluative relationships holding between outcomes, and it is odd to think that there are normative reasons for having preferences that misrepresent these relationships. Hence, with the exception of a few applications of EUT, Completeness will constitute an idealization of the descriptive kind. Second, Continuity constitutes a paradigmatic case of a descriptive idealization. It implies that preferences operate on an implausibly fine-grained level. This enables us to obtain a differentiable utility function and therefore affords us with mathematical convenience.

2.3. The Verdicts of Normative Models

In what follows, we will illustrate the guidance function of normative models by presenting examples of the kinds of normative verdicts that they are offering. As before, we shall focus on models based on EUT. These models are frequently regarded as being descriptively inaccurate models of decision making, that is, it is argued that they do not adequately capture people's choice behavior. Nonetheless, many authors think of EUT models as being normatively adequate models telling us how we should make decisions (see, for instance, Bleichrodt et al. 2001). In this regard, EUT models are frequently employed in order to guide decisions under risk (see Thoma 2019) or to overcome inconsistencies in our choice-behavior like, for instance, in the so-called Allais' paradox (see Mongin 2019). Hence—generally speaking—, EUT models are employed to derive verdicts that lend credence to certain courses of action.

To give a more vivid example of the verdicts EUT models are supposed to offer, consider that welfare economists are trying to develop techniques for deriving expected utility models of agents from alternative decision-theoretic models (like *cumulative prospect theory* or *rank dependent utility theory*), that are taken to be descriptively more adequate (see Bleichrodt et al. 2001; Tversky and Kahneman 1992). The aim of this exercise is to then use the EUT model to determine how the relevant agent *should* act (see Harrison and Ross 2017). In this regard, Bleichrodt et al. (2001) and Li et al. (2014) develop a framework that aims at guiding (medical) professionals who are supposed to act in the interest of their clients. The authors proceed as follows. In a first step, they present agents with a survey containing hypothetical choices to which they have to provide responses. In a second step, the authors try to derive a cumulative prospect theory model of the agents based on these responses; they use cumulative prospect theory here because they take it to provide *descriptively* adequate models of decision making. In a final step, the authors attempt to derive an EUT model of each agent from the cumulative prospect theory model. Crucially, they take the EUT model to be the *normatively* adequate model, whose verdicts can be used to determine how the agent *should* make certain decisions.

To illustrate this in more detail, Bleichrodt et al. (2001) employ this method in an experimental setup in which participants were asked to rank riskless treatments (i.e., treatments that would give them a certain amount of additional life years for sure) and risky treatments (i.e., treatments that could give them a certain amount of additional life years with a fixed probability but could also lead to them dying instantly). Based on the responses, the authors then estimated a cumulative prospect theory model of their participants and used it to derive an EUT model that is supposed to guide the participants. That is, the EUT model of a participant is supposed to tell her how she *should* have ranked the treatments, as opposed to how she actually ranked them, and also inform her about how she should make such comparisons in the future. Yet, during all of this, the authors offer only little insights into what it means for EUT models to be *normatively* adequate, let alone a discussion of how such models can offer guidance despite involving false descriptive statements, that is, descriptive idealizations.⁷

⁷Apart from EUT models—that are prominent in decision theory, applied ethics, and economics—we also find normative models in formal epistemology. On some interpretations, models of Bayesian belief updating—which, arguably, also involve a mixture of descriptive and normative idealizations—can be regarded as informing agents how they should update their beliefs (confer Colyvan 2013).

Let us sum up our characterization of normative models so far. Their main aim is to provide normative guidance toward agents. If they do so successfully, an agent would—other things being equal—be justified in following the model's verdicts. More specifically, the agent would act in accordance with a normative reason provided by the model. As noted by Colyvan (2013), normative models paradigmatically involve a mixture of normative and descriptive idealizations. We already anticipated that it is primarily the presence of descriptive idealizations that makes it difficult to see how normative models can successfully exert normative guidance. In particular, it is unclear in virtue of which property these models are apt to exert normative guidance. For the sake of brevity, let us call this property the *normative source* of the model. Hence, the puzzle about the guidance function boils down to the question of how we should conceive of the normative source of normative models.

Building on these insights, the aim of the next section is to highlight the absence of a systematic discussion about the normative source of normative models. Identifying this gap in the literature will pave the way for exploring two proposals for the normative source. We, thereby, offer some first steps toward outlining the success conditions of normative models.

Before proceeding, however, note that there are also other plausible ways to conceive of normative models. To illustrate, someone may want to use the term “normative model” to refer to models of a particular type of agent. For instance, consider models of expected utility maximizers or perfect Bayesian updaters. These particular types of agents, even if never fully instantiated, could be viewed as ideals. Consequently, we could think of the models that aim at representing these agents as descriptive models of an ideal. On the basis of such an understanding, one could try to argue that the puzzle this paper is concerned with does not exist: every model is descriptive, and normative models are just the subset of descriptive models that describe ideals.⁸

Yet, being a descriptive model of an ideal is not enough to be a normative model in our sense. To be a normative model in our sense, it is crucial that the model is taken to provide normative guidance, that is, that the model provides us with reasons for how we should act. However, if a model merely offers a description of how a certain type of agent looks like, this description does not, by itself, provide us with any normative reasons for how we should act. For example, Machiavelli's description of how an ideal prince would act does not, qua being a description, give us a reason

⁸We thank an anonymous reviewer for helping us to think through this issue more clearly.

to act in the same way as the ideal prince does. However, throughout this paper we maintain that there are normative models that can successfully exert normative guidance. The question we are concerned with is just in virtue of which property they can do so.

3. State of the Debate

As mentioned above, there is a rich literature attempting to reconstruct the success conditions of descriptive models, which has developed over the last 60 years quite significantly. Here, the main goal is to identify the property in virtue of which a model is supposed to represent its target-system and, thereby, achieves its descriptive aims. In the contemporary debate, philosophers have, for instance, pointed at features such as *similarity* holding between the model and its target (Weisberg 2013) or a certain kind of *isomorphism* (Bueno et al. 2002). Alternatively, they developed pluralistic accounts, for example, *feature-sharing* (Khosrowi 2018). Naturally, one would expect that there is a similar, parallel debate with respect to normative models and the property (i.e., their normative source) in virtue of which they meet their aim to successfully exert normative guidance.

In fact, within the respective scientific communities, particular normative models are often fiercely debated. For example, the normative appeal of EUT was the topic of a hot controversy between American economists like Samuelson, Savage, and Baumol in the 1950s (see Moscati 2018). The debate mainly focused on whether the above-mentioned Independence of Irrelevant Alternative assumption should be viewed as being normative.

However, only little work has been done on normative models as a general phenomenon and how they manage to successfully exert normative guidance. While recently there have been a few contributions on normative models, these only very tentatively address the issue at stake but rather take into focus other aspects of normative models. For instance, Colyvan (2013) mainly focuses on the question of how idealizations in normative models can be justified and argues that in this regard, there is little difference between normative and descriptive models. Moreover, Titelbaum (forthcoming) explores epistemic advantages and disadvantages of using normative models in order to be able to better compare them with alternative tools for normative inquiry. While he assumes that all of these tools can in principle be employed in tandem to elicit normative facts, he largely neglects that there is a specific puzzle about the normative source of normative models. Hence, neither of these authors is directly addressing what we take to be the core issue of normative modeling, that is, the question of how normative models successfully exert normative guidance.

Against this backdrop, we wish to explore two accounts of the normative source of normative models and critically review their respective strengths and weaknesses.

4. Normative Models and Normative Arguments

The first account states that normative models successfully exert normative guidance in an analogous way to how normative arguments successfully exert normative guidance. Consider the following normative argument establishing that you ought to regularly exercise.

- (1) You ought to live healthy.
- (2) Living healthy cannot be achieved without regularly exercising.
- (3) Therefore, you ought to regularly exercise.

The *normative* premise (1) expresses that there is a reason to live healthy, while it does not require that the agent is currently living healthy. The *descriptive* premise (2) makes a factual statement concerning a means necessary to achieve the end of living healthy. Importantly, for the argument to successfully exert normative guidance, its premises need to be sufficiently justified. However, the modes of justification for the normative and the descriptive premise of the argument differ. Normative premises need to be normatively justified, that is, it needs to be shown that there is (sufficient) normative reason for living up to the proposition contained in the ought-statement. In contrast, the mode of justification for descriptive premises in normative arguments is veridical, that is, there needs to be sufficient evidence indicating the truth of the premise. These considerations highlight that the normative source of normative arguments consists in the fact that a normative conclusion can be drawn on the basis of normative premises (for which we have normative justification) in conjunction with descriptive premises (for which we have support for their veridicality).⁹

Now, the first account of the normative source of normative models claims that it is analogous to the normative source of a normative argument: normative idealizations parallel normative premises, and descriptive idealizations (and possibly other descriptive assumptions) parallel descriptive premises. Importantly, both in the case of normative arguments and normative models,

⁹Note that our conception of what a normative argument is refers to what we take to be the standard understanding according to which arguments are *deductive*, that is, the premises deductively entail the conclusion.

the propositional content of normative premises/idealizations needs not be actualized, but it needs to be the case that there is (sufficient) normative reason to live up to them. Consequently, under the first account, the normative source of a normative model is constituted by the fact that certain normative verdicts follow from the model's assumptions of which at least one is a normative idealization. (This is important because, the thought goes, under this first account, descriptive idealizations on their own could not explain where normativity is ultimately coming from.) Note that while the first account does not claim that normative models *are* normative arguments, it basically denies that there is something particularly puzzling about the success conditions of normative models due to their resemblance to this very familiar normative device.

To illustrate the basic idea behind the first account, let us return to the example of EUT models. Under the first account, an agent would have reason to follow the verdicts of such models iff the agent has (sufficient) reasons to satisfy the normative idealizations of the EUT model (e.g., Transitivity) and it is the case that the other assumptions of the model pick out something true about the agent.

Having outlined this first account, we wish to argue that it faces serious challenges because it overstates the similarities between descriptive idealizations in normative models and descriptive premises in normative arguments. As explained in Section 2, in contrast to the descriptive premises of a normative argument, the descriptive idealizations of a model do not necessarily aim at correctly capturing a feature of the model's target. Instead, they are introduced to, for instance, reduce the complexity found in the target-system or to afford mathematical convenience. For example, recall that EUT models assume that preferences are continuous. This idealizing assumption is not motivated on any normative grounds, nor does it capture something that is likely to be correct about the preferences of many agents that are targeted by EUT models. Nonetheless, the idealization is necessary to derive certain results of such models.

We, therefore, hold that there is a stark disanalogy between normative models and normative arguments because descriptive premises in normative arguments require veridical justification, while this is not the case for descriptive idealizations in normative models. Thus, descriptive idealizations seem to play a different role in normative models than descriptive premises in normative arguments. While it is crucial for the latter to be veridically justified in order for the normative argument to exert normative guidance, the descriptive idealizations of normative models often (deliberately) misrepresent their target. To strengthen the disanalogy, just consider how easily defeasible the justifiability of descriptive idealizations would be *if one held that they require*

veridical justification. For instance, the implausibility of actually having infinitely fine-grained preferences would easily defeat the Continuity assumption of EUT models. Hence, we hold that the disanalogy puts the first account under serious pressure.^{10,11}

However, someone who wants to maintain the analogy between normative models and normative arguments with respect to their normative source may argue that descriptive idealizations are still linked to veridical justification: the descriptive idealizations of successful normative models are *close enough to the truth* or, put differently, they are *approximately true* descriptions of the agent that is targeted by the model. Given this kind of justification, real-world agents (this is, us) would nevertheless be successfully guided by the model and the first account introduced here would explain why this is the case; after all, the model's descriptive idealizations could, the thought goes, roughly play the role of true descriptive premises.

On the face of it, this is a powerful argument. An immediate reply would be to call into question that the descriptive idealizations of successful normative models are always approximately true. To illustrate that descriptive idealizations of successful models need not be approximately true, consider the assumption of an infinite population in game-theoretic models that are used in biology to study the influence of natural selection on the long-term evolution of phenotypic traits (see Potochnik 2017). It is not clear in which sense

¹⁰One may argue that there are some normative models that involve no descriptive idealizations whatsoever, but only true statements (and some normative idealizations). Because such models would involve no descriptive idealizations, they could be regarded as being analogous to normative arguments and the puzzle that we are concerned with in this paper would not apply to them. However, given that it is widely accepted that idealizations are “rampant and unchecked” (Potochnik 2017, 57-61; see also Weisberg 2007), we maintain that the disanalogy between normative models and normative arguments is widespread and, therefore, points at a serious puzzle for normative models.

¹¹Note that one could attempt to argue that premise (2) is a descriptive idealization as well. For instance, one could hold that it is a deliberative distortion of how things really are because it neglects the complex causal relationship between the effects of regularly exercising and living healthy. However, this misconstrues how normative arguments work. After all, premise (2) is not concerned with picking out the entire causal web surrounding living healthy, but just one necessary condition. In cases in which premise (2) is false (e.g., if some of the implicit assumptions necessary for it to be true do not hold), then it is, of course, false that the argument exerts normative guidance on the agent, i.e. the agent need not, then, regularly exercise. Therefore, the justification for premise (2) is clearly veridical. Thanks to an anonymous reviewer for pressing us on this issue.

this assumption should be regarded as approximately true. So why should we expect things to be different for normative models?

At this point, however, we want to grant that the objector is right in holding that the descriptive idealizations could sometimes be close enough to the truth and point out that even then, the reply is problematic. First, note that it is often assumed that a judgment about whether a model's idealizations are close enough to the truth depends *in part* on the aims the model sets out to serve (see Goodman 1972; Khosrowi 2018; Potochnik 2017). Applied to normative models, this means that whether a particular descriptive idealization of a normative model should be regarded as close enough to the truth can only be judged relative to whether it contributes to achieving its aim, that is, to provide normative guidance. Therefore, the state in which we could assess that the descriptive idealizations of a normative model are close enough to the truth would be a state in which we already have justified beliefs about whether its entailments are correct normative verdicts.¹²

This requirement, however, stands in conflict with the idea that the model's normative source is highly analogous to the one of normative arguments since in the case of a normative argument, justified beliefs about the derived normative verdicts are *not* required to assess the models guiding function. Quite to the contrary, normative arguments can successfully offer normative guidance precisely because we can rely on them independently of having any justified beliefs about their verdicts. To put the problem differently: the justification of the veridicality of a model's descriptive idealization seems to depend on already having justified beliefs in the verdicts of the model.

¹²At this point, one could argue that "providing normative guidance" is too vague for being a genuine aim of a model. Rather, the thought goes, models can only aim at guiding agents in specific contexts. Once we look at the specific context that a given normative model is applied to, we will see that taking into account the context provides us with reasons for holding that a specific descriptive idealization is true. For instance, if we know that our normative model aims at guiding an agent who only faces a very limited set of options, we may be warranted in holding that Completeness is satisfied, that is, true. However, while we do, of course, not deny that there can be contexts in which some of the assumptions that we call descriptive idealizations turn out to be true statements, this would not address the issue we are concerned with in this paper. The scope of application of models is seldomly restricted to the class of cases for which its assumptions are literally met (confer Weisberg 2007). Consequently, we do not think that it is a promising route to argue that, in the case of normative models, *all* of the descriptive idealizations are always literally met in *each* specific context in which the model provides normative guidance. Thanks to an anonymous reviewer for helping us to clarify this.

However, acquiring justification for those verdicts is why we appeal to the model in the first place. Consequently, requiring justified beliefs about the verdicts of the model would heavily undermine its guiding function. The only case in which the model could still provide some form of guidance would obtain if the agent was in the following epistemic state: She has sufficiently justified beliefs about the verdicts of the model in order to evaluate that its descriptive idealizations are approximately true, but the degree to which she is justified does not ultimately warrant to adopt the verdicts. (For if the agent was already warranted to adopt the verdicts, there would be no sense in which the model could still guide her.)

That the model could only provide guidance in the case just described would severely restrict the circumstances in which an agent would *need* the sort of normative guidance provided by a normative model. Moreover, it is also quite difficult to conceive of a case in which an agent is in this epistemic state. How should we conceive of an epistemic state in which the agent has strong enough justification in the model's verdicts to support that its descriptive idealizations should be regarded as approximately true, while the justification is at the same time too weak to ultimately adopt the verdicts of the model? Furthermore, even if we grant that agents can sometimes be in this peculiar epistemic state, it would require a lot of intricacy to argue that this epistemic state is present in all cases in which agents can be successfully guided by normative models.

Of course, without bothering too much about establishing the approximate truth of the descriptive idealizations, agents could simply make a leap of faith and trust that the descriptive idealizations are close enough to the truth. For instance, before relying on a verdict of a normative model, we could simply presuppose that the model's descriptive idealizations are close enough to the truth and, therefore, treat them as being harmless.¹³ However, holding that such a leap of faith is necessary in order to get into a position in which the model can successfully exert normative guidance is highly unsatisfactory. After all, we would lack grounds for relying on the model because we would have no concrete justification for the (approximate) veridicality of its descriptive premises, which is important under the first account.

At this point, one could try to argue that there is an easy fix to our worries. If we look at the case of descriptive models, we often see that people have sufficient grounds to justify the use of certain descriptive idealizations. For instance, one could justify the idealization that preferences are continuous when using EUT models as *predictive* models by arguing that making this

¹³Thanks to an anonymous reviewer for making this option more vivid to us.

idealization is harmless for the following reason: it has led to correct predictions in the past. This fact could be used to justify the gamble of also relying on this idealization for making predictions about other events. This gamble will pay off if the model offers a correct prediction for those events. However, while the fact that the model provided us with correct predictions in the past may justify the gamble of using its (descriptive) idealizations in the future, this in itself can hardly support that those idealizations were close enough to the truth. That is to say that following such a procedure would not give us the kind of veridical justification that is required by the first account. We have said before that whether a descriptive idealization should count as close enough to the truth partially depends on whether it facilitates the aims of the model. Yet, this does not entail that facilitating the aims of the model is sufficient for its assumptions to count as close enough to the truth. For instance, the ideal gas model in statistical mechanics introduces gas molecules as entities without extension. It is quite hard to see in which sense this should be regarded as an approximately true assumption. Nevertheless, according to Elgin (2017), the model achieves its epistemic aims. Consequently, it is important to bear in mind that the defender of the first account requires more from normative models than merely providing us with correct normative verdicts. She wants to explain why the model successfully exerts normative guidance by grounding the justification of its descriptive premises in their approximate truth. Hence, simply appealing to past successes of the model is of no help when it comes to defending the first account. What we need is approximate truth and for establishing this, past success is not enough.¹⁴

¹⁴One may be tempted to argue that we could do a kind of *robustness analysis* to check whether the descriptive idealizations of a normative model are undermining their guidance function (confer Odenbaugh and Alexandrova 2011). The general idea hereby would be to replace the descriptive idealizations with other assumptions and see whether the relevant verdicts remain the same. If this is the case, it could be argued that the presence of descriptive idealizations does not undermine the normative source of the model. However, note that some of the model's verdicts will certainly change in the process of substituting its descriptive idealizations. Consider, for instance, replacing the Continuity assumption of EUT with some discreteness assumption. It may then, for instance, no longer be the case that, assuming an agent prefers A to B to C, there is some mixture between A and C that the agent *should* treat as indifferent to B. Consequently, we need to have justified beliefs about what the relevant verdicts are that should *not* change, that is, the normatively correct verdicts, to execute the relevant robustness analysis. However, this would again put us in a spot where we require justified beliefs about what the correct verdicts in the domain of the model are to determine whether the model can successfully exert normative guidance.

Finally, someone who believes in the analogy between the normative sources of normative arguments and normative models has little to offer when it comes to explaining why we invoke descriptive idealizations in the first place. That is, from a normative standpoint it would appear undesirable to bring into play deliberate distortions of how the agent actually is. Of course, there are pragmatic rationales for invoking descriptive idealizations known from descriptive modeling such as affording us with mathematical convenience. However, given how a normative model would have to justify its descriptive idealizations under the first account (i.e., veridically), such rationales would fail to provide any warrant for the model's normative verdicts.

We take these considerations to constitute serious challenges for the first account. However, as a more radical way to save the first account, we briefly want to comment on the option of rejecting our core assumption that normative models contain a mixture of descriptive and normative idealizations. While we find this assumption convincing, one may try to argue that *all* idealizations are normative, that is, that the alleged descriptive idealizations (such as Continuity in the case of EUT) are normative idealizations as well. This could save the first account by simply eliminating descriptive idealizations and, therefore, the disanalogy between normative models and normative arguments. For now, we take this to be a too drastic departure from how such idealizations are usually understood and leave it to further research whether anything like this can be established. Especially, note that the challenge is not to show that it would be “nice” to have, say, continuous preferences, but to establish that an agent has a genuine normative reason for having those. However, if this could be established (notably, for all relevant normative models), the puzzle about the normative guidance function of normative models would simply disappear since all of the model's idealizations would just encode normative reasons the agent has anyway. Note, however, that departing from Colyvan's (2013) argument that normative models involve a mixture between normative and descriptive assumptions could also lead to seriously undermine the first account; for, on the other extreme, one may try to defend that all idealizations of a normative model are descriptive. Yet, if this were the case it would now be even more mysterious how normative models are supposed to normatively guide us if we adhere to the first account.

In sum, the first account, which seeks to identify an analogy between the normative source of normative models and the one of normative arguments, faces serious challenges. Nevertheless, one may still wish to argue for this account by defending that normative models' descriptive idealizations are, indeed, close enough to the truth. This endeavor might be difficult because of the obstacles laid out above, but we leave it to further inquiry whether there

is a way to successfully carry out this strategy. In the meantime, we maintain that, in contrast to the normative source of a normative argument, there is something puzzling about the normative source of normative models that makes it unclear how and when they can provide normative guidance. Given all of this, we now explore an alternative account that has the potential to elucidate the success conditions of normative models, while sidestepping the problems that descriptive idealizations pose for the normative source of normative models under the first account.

5. Normative Models as Extending Normative Justification

5.1. A First Alternative Attempt: An Extrinsic Normative Source

While the first account for explaining the normative source of normative models attributes a direct justificatory function to individual elements of normative models by grounding their normative source in their assumptions, one may be tempted to pursue the opposite route: The normative source lies entirely outside of the normative model, that is, the normative verdicts of the model are justified entirely by extrinsic considerations. That is, one could hold that the verdicts are justified by some model-independent normative reasoning. For example, one may have considered judgments that support the model's verdicts (see Rawls 1971).¹⁵ In the following, we first introduce an attempt at developing such an account and explain why, even though it may be *prima facie* plausible to conceive of normative models in that way, it faces difficulties in accounting for their guiding function. We will then introduce a more promising alternative that intricately connects the guiding function of normative models to their ability to extend normative justification.¹⁶

¹⁵In the rest of the paper, we shall—for the sake of simplicity—sometimes assume that such judgments can ground normative justification even though we are not committed to that.

¹⁶Note that one could argue that we are too quick in abandoning the idea that the normative source of a normative model is grounded in its assumptions. In particular, one may hold that there is a way of grounding the normative source of normative models in its assumptions that does not depend on their analogy to normative arguments. On such a view, there might be a different kind of justification for descriptive idealizations that does not pertain to their veridicality. However, one would then have to show that this kind of justification (whatever it may be) can underpin the model's ability to exert normative guidance. As we already pointed out, there are certainly *pragmatic* rationales for descriptive idealizations. Yet, we do not see how such pragmatic rationales

A first attempt at developing an alternative to the first account may hold that a normative model counts as normatively adequate because it reproduces independently justified normative propositions; yet, the normative model does not play any justificatory role itself. Under this account, the normative model could only be said to exert normative guidance insofar as the model would be a heuristic for directing us at normative verdicts that are independently justified. For instance, one may hold that the model is able to exert normative guidance by directing us toward our considered judgments.

For illustration, consider a normative model that provides you with a verdict concerning how you should choose between different risky medical procedures (recall our example from Section 2.3). If this verdict matches your considered judgments (assuming that they can ground normative justification), it could be said that you have a reason to follow the verdict. Yet, under the view in question, the normative model would merely be capable of generating the respective verdict without providing any further reason to accept it. In that way, the kind of normative guidance that normative models could provide would be dispensable. After all, we could obtain normative guidance directly from the sources that justify the normative verdicts in the first place, and the model would not do any normative work. Since the view in question would, thereby, threaten to trivialize the role of normative models, we do not want to further expand on it here. Normative models do not seem so easily dispensable concerning their normative guidance function—an intuition that apparently underlies the appreciation of normative models among many practitioners of, for instance, decision theory.

5.2. A Promising Alternative: Extending Normative Justification

In the light of these considerations, we now wish to explicate a genuine alternative to the first account outlined in Section 4. This alternative grants the

can ground a model's normative guidance function. Nevertheless, a defender of the view in question could still try to point at another non-veridical kind justification for descriptive idealizations that would render them part of the normative source of normative models. Yet, by alluding to a non-veridical kind of justification for descriptive idealizations, we would have to give up on the analogy to normative arguments and, thereby, on explaining the normative source of normative models by showing their similarity to this familiar normative device. As of now, we are unaware of how a plausible story along the lines just sketched would look like, and, therefore, leave the idea that the normative source of a normative argument is entirely grounded in its assumptions behind from here on.

importance of independently justified verdicts but does not trivialize the model's guidance function. Generally speaking, this second account grounds the normative source of a normative model in two things: first, some intrinsic feature of the model and, second, an array of normative verdicts, whose justification is extrinsic to the model, that is, established by model-independent normative reasoning. Thus, for understanding the kind of normative guidance provided by a normative model, both of these features are relevant.

Let us flesh out this account in more detail. According to it, the way in which normative models exert normative guidance is by means of extending normative justification to cases of normative uncertainty (i.e., a situation in which we are unsure about which action we should perform).¹⁷ Let us call this the model's *extending function*. This function consists of two elements, which we shall elaborate on in turn. First, normative models allow us to summarize normative verdicts that are justified independently of the model. They do so by means of a (typically) sparse set of idealizing assumptions that entail these verdicts. Thus, normative models are economical tools for capturing these verdicts.

In order to get a better grip on this, a useful analogy from descriptive modeling can be provided by *revealed preference theory* in the study of consumer choice. The main idea here is that if people's choices exhibit certain characteristics (e.g., satisfying the so-called weak axiom), then there exists a utility function from which all of these choices can be derived. Therefore, it is often said that utility functions in consumer choice theory are appropriately viewed as concise summaries of large sets of choice behavior (Clarke 2016). By providing these summaries, utility functions allow us to reveal a pattern in these choices. Similarly, we hold that normative models can play the role of offering concise summaries of our normative verdicts that allow us to reveal patterns in those verdicts. They can do so in virtue of entailing a large array of normative verdicts on the basis of a set of a few (idealizing) assumptions. Yet, as will become clear shortly, this deductive relationship does not—in contrast to the first account—exhaust the normative source of the normative model.

The second element of the extending function consists in the fact that normative models allow us to project the identified patterns onto novel situations. In that way, normative models can increase the number of justified normative

¹⁷Note that our use of the term “normative uncertainty” differs from how it is used in a contemporary debate that revolves around the issue of how we should make decisions given that we are often justifiably uncertain about what the correct normative or moral theory is (see, for instance, Bykvist et al. 2020).

verdicts over and above the ones that we obtain from model-independent sources of normative justification. To pick up on the analogy with consumer choice theory, utility functions are here used to project patterns in choice behavior onto novel situations (i.e., to predict future choice behavior). Similarly, we hold that the patterns identified by the summaries offered by normative models can be projected onto cases of normative uncertainty.

To illustrate both parts of the extending function, suppose that we are in a case of normative uncertainty, that is, we are unsure what we should do. In such a situation, a normative model can help in the following way. After it enabled us to summarize a large class of justified normative verdicts, it may then also be used to project the pattern that was revealed by the model's summarizing function onto cases of normative uncertainty for which we previously lacked firm verdicts. Our main claim is now that, thereby, a normative model connects these cases of normative uncertainty to others about which we already have justified verdicts and, in this sense, *extends* justification to the cases of normative uncertainty. In virtue of extending justification, the model can *successfully* exert normative guidance. If this is correct, a normative model can enlarge the number of situations for which we have a justified verdict about what we should do compared to the state prior to appealing to the model. Under the second account, which we will also call the *extending account*, normative models can, therefore, offer a non-trivial form of guidance.

On an abstract level, normative models would follow a certain schema when successfully exerting guidance. Let p_1, \dots, p_n denote an array of normative verdicts for which we have a (sufficient) normative justification independent from the model. A normative model reproduces these verdicts by means of a sparse set of idealizing assumptions and, thereby, reveals a projectable pattern in those verdicts. Now, suppose we are unsure whether to accept some normative verdict q , and that q follows from the model. Then, according to the envisaged account, there is a normative reason to act in accordance with q because it fits well with the pattern of our prior, justified normative verdicts that is encoded by the model.

What does this mean for the normative source of normative models, that is, the property in virtue of which the model is able to *successfully* exert normative guidance? First, for the model to be able to successfully exert normative guidance, there needs to be a substantial body of independently justified normative verdicts. Second, the model has to be such that it enables a concise summary of those verdicts that reveals patterns that can be then projected onto cases of normative uncertainty. All of this reflects the general shape of the account that we anticipated above: on the one hand, the normative source is partially grounded in the intrinsic feature of the model to identify and

project a pattern in our verdicts. On the other hand, the reason why these prior verdicts are justified is extrinsic to the model.

To better appreciate the second account, consider the following example. Think of a student of decision theory, who, so far, found herself to strongly agree with the verdicts of EUT models presented to her. Now, imagine that she participates in an experimental setup similar to the one of Bleichrodt et al. (2001) that we mentioned in Section 2. The student is asked to rank several choice options. Based on her responses, the experimenters then derive an EUT model of the student. Given that the student so far agreed with the verdicts of EUT, it is also quite likely that many verdicts of this EUT model will match her considered judgments. However, there may also be some cases within the domain for which the model issues verdicts about which the student lacks considered judgments. Now, according to the extending account, the model provides her with some reason to follow its verdicts also in those cases because these verdicts fit into the patterns of the verdicts supported by her considered judgments. If this were not the case, that is, if there was not a substantial body of prior, considered judgments that match some of the model's verdicts, the model would fail at normatively guiding the student.

To put it differently, the student has reason to accept certain verdicts of the model since it already captures a lot of verdicts that match her considered judgments. Consequently, she can successfully rely on the relevant model to obtain guidance for situations in which she lacks considered judgments about what she ought to do. In terms of the extending account, the student employs the EUT model as a means to extend the justification from cases where she had firm judgments to this new case. Again, what is crucial for this story is that if the student's considered judgments would not match many of the other verdicts of the EUT model, the model could not successfully guide her.

The example illustrates that, under the account in question, a normative model exerts a rather indirect form of normative guidance. This is because the normative model does not provide, say, a direct argument for why to accept certain verdicts. Rather, it provides a reason to accept a verdict because it would fit into the pattern of justified normative verdicts that the normative model is unraveling. We think of the reason generated by the extending function to be relatively weak compared to a reason provided by, say, a normative argument. Consequently, the extending account seems to identify a relatively weak kind of normative guidance.

Nevertheless, we take this account to offer a promising answer to the puzzle of how normative models can successfully exert normative guidance. While the extending account only grants normative models the ability to successfully exert a relatively weak kind of guidance, it, thereby, nonetheless

attributes a genuine guidance function to normative models. Of course, some proponents of normative models may wish to establish that normative models can provide us with a stronger form of guidance. They could, therefore, argue that the extending account fails to provide a satisfying answer to the puzzle we introduced here because any satisfying answer would have to grant a stronger guidance function to normative models. Our reply to this is that how strong the guidance function of normative models ultimately is depends on their normative source. In this regard, we invite those who want to establish a stronger guidance function for normative models to come up with an alternative account of their normative source or to defend the account outlined in Section 4. Moreover, it is, to our minds, not puzzling at all that people, in the absence of an explicit account of the success conditions of normative models, overestimate the kind of normative guidance that these models can offer. Nevertheless, we hold that it would be very puzzling if normative models were taken to successfully exert *some* normative guidance while they have no normative source at all. In light of this, we hold that the extending account offers a promising answer to the puzzle, not least because it can sidestep the problems that arose for the first account due to the presence of descriptive idealizations.

Let us elaborate on this point and outline why the extending account can establish that the descriptive idealizations of normative models are unproblematic. On the one hand, it can elucidate why we introduce these idealizations in the first place: they help to reproduce a vast variety of verdicts by means of a sparse set of idealizing assumptions. On the other hand, descriptive idealizations can be reconciled with the normative guidance function of normative models because they are not employed like in a *deductive* normative argument, but are rather part of a more indirect kind of normative justification. In this regard, it is important to note that the extending account could also explain why normative models can offer normative guidance if they were to only involve descriptive idealizations since the extending function does not depend on the presence of normative idealizations.¹⁸ Hence, even if one were to reject the force of, say, money-pump arguments in favor of the Transitivity assumption of EUT (which is not an implausible position, see Aldred 2003), the extending account could still explain how EUT models can offer normative guidance.

¹⁸Note that under the extending account, descriptive *and* normative idealizations have the same function, namely, to capture a projectable pattern in our justified normative verdicts. While it is plausible that capturing such a pattern will involve normative idealizations, their presence is not required for the extending account.

Apart from putting forward an answer to the puzzle we are concerned with in this paper, the extending account also renders normative models as quite interesting normative devices that can enrich our toolbox for normative justification. By extending normative justification, normative models sit next to a device that is commonplace in normative theorizing (and receives plenty of attention in the current meta-normative literature): the method of reflective equilibrium (Rawls 1971, 46-53; Scanlon 2014, Chapter 4). What normative models under the extending account share with the method of reflective equilibrium is that the latter can also justify verdicts by means of extending normative justification (see McGrath 2019, 19): if we obtained a systematic body of principles and verdicts via the method of reflective equilibrium, then we have reason to accept verdicts that follow from the general principles since they would, thus, fit well with that systematic body. Hence, normative justification is—in this sense—extended to the new verdict.

Importantly, however, normative models operate differently from reflective equilibrium in several respects. First off, normative models contain descriptive idealizations, which are not supposed to be normative principles or normative verdicts. Contra that, the method of reflective equilibrium deals with principles and verdicts all of which are supposed to express normative content. Moreover, the function of a normative model's idealizations is crucially different from the function of principles in the method of reflective equilibrium: the normative justifiability of these principles is itself at stake when applying the method of reflective equilibrium and the principles are also supposed to provide some normative justification for accepting less general normative verdicts. Something parallel is not true about normative models under the extending account: their idealizations do not stand in a direct justificatory relationship to the verdicts nor do the verdicts to the idealizations. Rather, according to the extending account, their function is merely to reproduce a pattern in our normative verdicts, which can be then projected onto cases of normative uncertainty. Hence, reflective equilibrium offers us a more direct route of justification than the extending function of normative models.

Consequently, the extending account would also suggest an important limitation of normative models. To exert normative guidance, it is required that the agent appealing to the normative model already possesses a certain amount of independently justified normative verdicts that can be reproduced by the model. If this is not the case, there is no sense in trying to use this model to project the pattern in those verdicts to novel situations. In this regard, and in contrast to what other authors have mentioned (e.g., Titelbaum forthcoming), normative models would not be on par with other tools for normative inquiry like reflective equilibrium, as they would require these

tools to establish the normative verdicts on which the extending function can operate in the first place.

As a final note, we hold that interesting questions for further research crop up if we accept the extending account of the guidance function of normative models. The first question revolves around the strength of the extending function: what is the degree of normative justification the model needs to provide such that it is reasonable to accept a verdict we were uncertain about before? Probably, important parameters here not only include the number of independently justified verdicts that are captured by the model, but also their strength of justification. Second, note that we have restricted us in this paper to the relatively simple case of extending normative justification to cases of normative uncertainty. Yet, one might inquire whether the extending function provides further normative support regarding verdicts we already hold. Moreover, there might even be situations in which the extending function can imply that we should override some of our normative verdicts. Finally, another interesting question is how we should deal with situations in which different normative models reveal different patterns in our normative verdicts that lead to conflicting results in cases of normative uncertainty.

6. Conclusion

In this paper, we tried to highlight the lack of a systematic discussion of the success conditions of normative models. We argued that the absence of such a debate is surprising because there is an extensive discussion of the success conditions of descriptive models, that is, in virtue of which properties they achieve an adequate representation of their target-system. Consequently, one would expect a similar debate that pertains to the success conditions of normative models, that is, in virtue of which properties they exert normative guidance. Of course, one explanation for why there is no debate about the success conditions of normative models is that they exert their normative guidance in the same way as more familiar normative devices. Indeed, on a first look, it may be tempting to hold that normative models successfully exert guidance similar to how normative arguments do. However, we argued that there are serious challenges to this proposal because it underestimates the role of descriptive idealizations in normative models. We, therefore, offered an alternative account of how normative models exert normative guidance that can sidestep the challenges faced by the first account.

According to the second account, the normative guidance function of normative models consists in their ability to extend normative justification from cases for which we already have justified verdicts to cases in which we lack such verdicts. Normative models manage to do this because they

can capture and reveal patterns in our already justified verdicts. On the basis of that, normative models can yield novel verdicts and reasons to accept these verdicts. However, we have also pointed out that, under the extending account, the normative guidance function of normative models may turn out to be relatively weak. Nevertheless, the account identifies a plausible sense in which normative models can successfully exert normative guidance. Hence, we consider the extending account a promising, albeit minimalist proposal for how we should conceive of the normative guidance function of normative models.

While we are uncertain about whether any stronger guidance function can be established, we invite others to come up with a more ambitious proposal like defending the first account discussed here. Other interesting questions include whether, upon further inquiry, different normative models will turn out to have different normative sources, or whether one and the same model can have multiple normative sources. Given that so much attention has been devoted to figuring out the success conditions of descriptive models, the investigation of the success conditions of normative models certainly merits more attention. In this spirit, we hope that this paper gets the ball rolling on what we take to be the core philosophical issue of normative modeling.

Acknowledgments

For their helpful comments on earlier versions of this article, we would like to thank Anna Alexandrova, Philip Fox, Moritz Graefrath, Kareem Khalifa, Cristian Larroulet Philippi, Mark Risjord, Thomas Schmidt, Bele Wollesen, Kevin Zollman, as well as two anonymous referees. We also benefitted from useful feedback from audiences at the 2020 Philosophy of Social Science Roundtable, Anna Alexandrova's PhD Reading Group at the University of Cambridge, and Thomas Schmidt's Ethics Colloquium at Humboldt-Universität zu Berlin.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Aldred, Jonathan. 2003. "The Money Pump Revisited." *Risk Decision and Policy* 8 (1): 59-76.

- Alexandrova, Anna. 2018. *A Philosophy for the Science of Well-Being*. Oxford: Oxford University Press.
- Bleichrodt, Han, Jose Luis Pinto, and Peter P. Wakker. 2001. "Making Descriptive Use of Prospect Theory to Improve The Prescriptive Use of Expected Utility." *Management Science* 47 (11): 1498-514.
- Broome, John. 2013. *Rationality Through Reasoning*. Chichester: Wiley-Blackwell.
- Bruni, Luigino, and Robert Sugden. 2007. "The Road Not Taken: How Psychology Was Removed from Economics, and How It Might Be Brought Back." *Economic Journal* 117 (516): 146-73.
- Buchak, Lara Marie. 2013. *Risk and Rationality*. Oxford: Oxford University Press.
- Bueno, Otávio, Steven French, and James Ladyman. 2002. "On Representing the Relationship between the Mathematical and the Empirical." *Philosophy of Science* 69 (3): 497-18.
- Bykvist, Krister, William MacAskill, and Toby Ord. 2020. *Moral Uncertainty*. Oxford: Oxford University Press.
- Chang, Ruth. 1997. *Incommensurability, Incomparability, and Practical Reason*. Cambridge: Harvard University Press.
- Clarke, Christopher. 2016. "Preferences and Positivist Methodology in Economics." *Philosophy of Science* 83 (2): 192-212.
- Colyvan, Mark. 2013. "Idealisations in Normative Models." *Synthese* 190 (8): 1337-350.
- Dancy, Jonathan. 2004. "Enticing Reasons." In *Reason and Value: Themes From the Moral Philosophy of Joseph Raz*, edited by R. J. Wallace, P. Pettit, S. Scheffler, and M. Smith. Oxford: Clarendon Press.
- Elgin, Catherine Z. 1983. *With Reference to Reference*. Indianapolis: Hackett.
- Elgin, Catherine Z. 2017. *True Enough*. Cambridge: MIT Press.
- Fox, Philip. 2019. "Revisiting the Argument from Action Guidance." *Journal of Ethics and Social Philosophy* 15 (3): 222-54.
- Frigg, Roman, and James Nguyen. 2016. "The Fiction View of Models Reloaded." *The Monist* 99 (3): 225-42.
- Goodman, Nelson. 1972. *Problems and Projects*. Indianapolis: Bobbs-Merrill.
- Harrison, Glenn W, and Don Ross. 2017. "The Empirical Adequacy of Cumulative Prospect Theory and Its Implications for Normative Assessment." *Journal of Economic Methodology* 24 (2): 150-65.
- Hersch, Gil. 2020. "No Theory-Free Lunches in Well-Being Policy." *The Philosophical Quarterly* 70 (278): 43-64.
- Hesse, Mary B. 1963. *Models and Analogies in Science*. London: Sheed and Ward.
- Khosrowi, Donal. 2018. "Getting Serious about Shared Features." *British Journal for the Philosophy of Science*. Published electronically June 28 2018. doi:10.1093/bjps/axy029.
- Li, Chen, Zhihua Li, and Peter P. Wakker. 2014. "If Nudge Cannot Be Applied: A Litmus Test of the Readers' Stance on Paternalism." *Theory and Decision* 76 (3): 297-315.
- Maguire, Barry. 2016. "The Value-Based Theory of Reasons." *Ergo* 3 (9): 233-62.

- Mäki, Uskali. 2009. "MISSing the World. Models as Isolations and Credible Surrogate Systems." *Erkenntnis* 70 (1): 29-43.
- McGrath, Sarah. 2019. *Moral Knowledge*. Oxford: Oxford University Press.
- Mongin, Philippe. 2019. "The Allais Paradox: What It Became, What It Really Was, What It Now Suggests to Us." *Economics and Philosophy* 35 (3): 423-59.
- Morgan, Margaret, and Mary Morrison. 1999. *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press.
- Moscati, Ivan. 2018. *Measuring Utility: From the Marginal Revolution to Behavioral Economics*. New York: Oxford University Press.
- Odenbaugh, Jay, and Anna Alexandrova. 2011. "Buyer Beware: Robustness Analyses in Economics and Biology." *Biology & Philosophy* 26 (5): 757-71.
- Potochnik, Angela. 2017. *Idealization and the Aims of Science*. Chicago: University of Chicago Press.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge: Harvard University Press.
- Scanlon, Thomas M. 2014. *Being Realistic about Reasons*. Oxford: Oxford University Press.
- Schroeder, Mark. 2007. *Slaves of the Passions*. Oxford: Oxford University Press.
- Thoma, Johanna. 2019. "Decision Theory." In *The Open Handbook of Formal Epistemology*, edited by R. Pettigrew and J. Weisberg. London: The PhilPapers Foundation.
- Titelbaum, Michael. forthcoming. "Normative Modeling." In *Methods in Analytic Philosophy: A Contemporary Reader*, edited by J. Horvath. London: Bloomsbury Academic Press.
- Tversky, Amos. 1969. "Intransitivity of Preferences." *Psychological Review* 76 (1): 31-48.
- Tversky, Amos, and Daniel Kahneman. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47 (2): 263-91.
- Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty* 5 (4): 297-323.
- Von Neumann, John and Morgenstern Oskar. 1944. *Theory of Games and Economic Behaviour*. New Jersey: Princeton University Press.
- Weisberg, Michael. 2007. "Three Kinds of Idealization." *The Journal of Philosophy* 104 (12): 639-59.
- Weisberg, Michael. 2013. *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.

Author Biographies

Lukas Beck is a PhD student in the Department of History and Philosophy of Science at the University of Cambridge. His research interests include philosophy of economics, philosophy and public policy, and the philosophy of the social sciences.

Marcel Jahn is a PhD student in the Department of Philosophy at the Humboldt-Universität zu Berlin. His research interests include moral philosophy, the foundations of normativity, and the philosophy of the social sciences.