**COMPUTATIONAL MODELLING IN PHILOSOPHY**

# Argumentative landscapes: the function of models in social epistemology

**N. Emrah Aydinonat[1]** · **Samuli Reijula[2]** · **Petri Ylikoski[3]**

## Abstract

We argue that the appraisal of models in social epistemology requires conceiving of them as argumentative devices, taking into account the argumentative context and adopting a family-of-models perspective. We draw up such an account and show how it makes it easier to see the value and limits of the use of models in social epistemology. To illustrate our points, we document and explicate the argumentative role of epistemic landscape models in social epistemology and highlight their limitations. We also claim that our account could be fruitfully used in appraising other models in philosophy and science.

**Keywords** Social epistemology · Modelling in philosophy · Epistemic landscape models · Argument view · Epistemic value

## 1 Introduction

Philosophers employ a wide variety of argumentative resources to support their claims, including appeals to intuition, authoritative testimony, common sense, case studies, formal derivations and even experiments. The usefulness of these resources must be evaluated with respect to their contribution to the strength of a philosophical argument. In the present paper we suggest that the same principle applies to the use of mathematical and computational models in philosophy. We argue that to understand the value—and limits—of models in social epistemology, one needs to analyse them in the context of the arguments they are used to support.

✉ N. Emrah Aydinonat
  emrah.aydinonat@helsinki.fi

1 Practical Philosophy, Faculty of Social Sciences, University of Helsinki, P.O. Box 24, 00014 Helsinki, Finland

2 Theoretical Philosophy, Faculty of Arts, University of Helsinki, P.O. Box 24, 00014 Helsinki, Finland

3 Sociology, Faculty of Social Sciences, University of Helsinki, P.O. Box 18, 00014 Helsinki, Finland

🍦 Springer

We further argue that models in social epistemology function as argumentative devices. When philosophers build and use mathematical and computational models, their conclusions often go beyond what can be supported by the premises of their model. Models are often used to support a broader philosophical claim, and the model functions as one of the many premises of the argument. Thus, even if one conceives of a model as an argument in itself (Beisbart 2012), there may well be a gap between what it can establish (i.e., the model argument) by itself and what the philosopher wants to conclude (i.e., the philosophical argument). This gap is filled with argumentative steps, many of which remain implicit. It is crucial to understand these steps to make sense of how models are used in philosophical argumentation. Our aim in this paper is to make sense of the functioning of models in this context. Using epistemic landscape (EL) models in social epistemology as our example, we illustrate how an EL model operates in a philosophical argument. As we will show, being explicit about the several steps involved helps in terms of seeing through the argument and assessing its strengths and weaknesses.

Second, we describe how changes introduced into an existing model could be understood as argumentative moves. Philosophers who use models present their arguments in the context of existing arguments, as amendments, criticisms or plausible alternatives. It is necessary to understand this argumentative context to make sense of the model's contribution. Taking into account this broader context shows how the success of a model (or a modelling framework) does not depend solely on its intrinsic properties. The epistemic contribution of a model typically becomes clear after a collective process of modifying, reinterpreting and criticising the model and its suggested implications. For example, as we will show, not only did Weisberg and Muldoon's (2009) epistemic landscape model suffer from deficiencies such as a lack of representational adequacy and derivational robustness,[1] it also contained programming errors. Nevertheless, it became one of the most well-known models in social epistemology. If one ignores the dynamics of the later dispute and collaboration, it is hard to understand its success. More importantly, taking the broader argumentative context into account sheds light on how the meaning and value of a model may be established long after its original publication.

Like many other models in social epistemology (e.g., Kitcher 1990, 1993; Strevens 2003; Hegselman and Krause 2009), EL models are highly idealised and sensitive (i.e., not robust) to changes in their structural assumptions or parameters. Until now, most of the discussion on the value of EL models has focused mainly on the model-target dyad (to use Knuuttila's term 2009a, b) and robustness. We suggest that such a perspective does not suffice, and argue that to see the value and limits of EL models one must take into account the fact that they function as argumentative devices. The representational adequacy of a model can only be evaluated if one understands the argumentative goals it is intended to serve, and knowledge of its robustness may well only emerge over time as others modify the assumptions and the implementation of the original model. Furthermore, one can assess the credibility of a model without engaging in a pairwise model-target comparison—which would be very difficult in the

---

[1] In the present paper, we use 'robustness' to indicate 'derivational robustness', which concerns the sensitivity of a model's results to the alterations of its assumptions (Woodward 2006). In Sect. 5, we briefly discuss how derivational robustness analysis relates to a broader notion of robustness.

case of highly abstract models—by making piecemeal assessments of the plausibility of individual assumptions relative to the argumentative goals.

The paper proceeds as follows. In the next section we briefly introduce the Weisberg–Muldoon model and show that a lack of representational adequacy and robustness are the two main concerns about the use of abstract models in social epistemology. These are reasonable concerns. We argue, however, that focusing solely on representation and robustness is not appropriate because it does not take the argumentative goals and context into account. We also suggest that consideration of the argumentative context, goals and moves elucidates the intended and actual contribution of a model. Section 3 outlines our account of models as argumentative devices. We claim that understanding how a model is used in a philosophical argument requires disentangling several layers of argumentation. By way of illustration, in Sect. 3.1 we disentangle three such layers–the philosophical argument, the conceptual model and the computational model—in Weisberg and Muldoon's (2009) argument and show how their main conclusions are supported. In Sect. 3.2 we examine their argument in detail, discuss how the credibility of the model can be assessed, and illustrate the gaps and problems in their argumentation. We further argue that to grasp the value-added of an abstract philosophical model, one must question how it supports claims about difference making. We discuss how difference making can be established and introduce the idea that epistemic benefits from modelling are commonly realised as a consequence of several modelling attempts.

With a view to analysing how the process of collective exploration ultimately determines the value of a model, in Sect. 4 we introduce the notions of argumentative goal and argumentative move. To illustrate the usefulness of these notions and our approach, we discuss what we call first- and second-generation EL models. In Sect. 5, we broaden our view from individual models to model families to show that such a perspective provides a better understanding of how EL models establish difference makers, and how the discovery of both robust and non-robust results might serve useful purposes. We also show how the argumentative force of the Weisberg–Muldoon model gradually emerges in the collective work of people who refined, criticised and explored the EL framework. Finally, we argue that analysing a model (e.g., the Weisberg–Muldoon model) in isolation from the argumentative context and other related models cannot encapsulate the understanding created by this family of models. The final section of the paper presents our concluding observations.

## 2 The research topic as an epistemic landscape

Weisberg and Muldoon (2009) developed an agent-based model of a population of scientists to argue for the epistemic usefulness of cognitive diversity in scientific communities. Adapting a biological modelling paradigm introduced by Wright (1932), the model portrays scientists as agents foraging on a landscape that stands for a scientific research topic. Each patch on the landscape represents a particular research approach, and the elevation of the patch corresponds to the epistemic significance of the approach. The population of agents faces the task of finding approaches of non-zero epistemic significance. There are three kinds of agents: controls, followers and

mavericks. Controls do not pay attention to which approaches have been adopted by others; followers have a conservative research strategy and tend to adopt already tried-out research approaches; and mavericks prefer to explore new approaches. Having examined populations of agents consisting of different mixes of the three kinds, Weisberg and Muldoon suggest that the existence of mavericks in a population enhances its capacity for making epistemic progress. They contend that maverick agents act as pathbreakers who help follower agents to find research approaches of high epistemic significance. One striking outcome of the model is that best epistemic performance is achieved by a scientific community consisting solely of maverick agents.

The Weisberg–Muldoon model has been followed by a series of models inspired by the original (see Reijula and Kuorikoski 2019 for a review). However, further scrutiny revealed a number of shortcomings in Weisberg and Muldoon's model that shed doubt on its epistemic value. For example, Alexander et al. (2015) showed that Weisberg and Muldoon's conclusions were undermined by an implementation error. Correcting this error and adding a slight cost of exploration to the model leads to a situation in which the best epistemic performance is no longer achieved by a 100% maverick population, it is achieved by a polymorphic population (Pöyhönen 2017). The model has also been criticised on various other grounds. Thoma (2015), for example, showed that alternative search rules that appear to be just as compatible with the behaviour of real scientists as those suggested by Weisberg and Muldoon lead to clearly different collective outcomes. Furthermore, the model's depiction of an epistemic landscape has been criticised as being too simple to capture actual scientific-research domains (Alexander et al. 2015; Pöyhönen 2017). Martini and Pinto (2017) question the epistemic value of models of the social organisation of science, including the Weisberg–Muldoon model, because their representational adequacy is not empirically established. Finally, explorations of the model show that its results are not robust (Thoma 2015; Alexander et al. 2015; Reijula and Kuorikoski 2019; Pöyhönen 2017; Pinto and Pinto 2018). In sum, the main concerns about the Weisberg–Muldoon model are the following: conceptual problems and errors, a lack of representational adequacy and non-robustness.

These concerns reflect more general worries about the use of models in social epistemology. Bedessem (2019, p. 3) argues that models of the division of cognitive labour suffer from a "fundamental lack of clarity about the exact object which is divided", and an ambiguity about what the division of labour is in reality. He also questions the representational adequacy of models in social epistemology because they ignore essential components of scientific practice and progress: in ignoring the fact that scientific fields are hierarchical and interconnected complexes, such models overlook the context dependency of epistemic benefits deriving from cognitive diversity and pluralism. Rosenstock et al. (2017) argue in their analysis of epistemic network models that because most of them are not representationally adequate, what they reveal about real epistemic networks is limited:

> We do not have a good sense of which real world communities are well represented by which epistemic network models. This is because these models are highly simplified. They abstract away from many relevant details of such communities. (Rosenstock et al. 2017, p. 250)

Thicke ([2019](#)) puts representational adequacy at the centre of his analysis of the epistemic value of formal models of science and expresses his concerns about their theoretical and predictive value. He argues that "the plausibility established by most formal models of science is very weak; while there might be some similarities between the organization of scientific communities and the structure of these models, it is often a very distant sort of similarity" (Thicke [2019](#), p. 17). Thicke's other concern, robustness, is also shared by other authors (e.g., Rosenstock et al. [2017](#); Frey and Šešelja [2018](#)).

> When epistemic network effects are highly robust, it makes sense to take them more seriously as important findings for real world communities. (Rosenstock et al. [2017](#), p. 250)

Philosophers of science tend to use representational adequacy and robustness to evaluate the epistemic value of a model. The intuitions behind these criteria are straightforward. First, although philosophers do not agree on what exactly representational adequacy is, they do generally agree that to be able to establish a link between the model and its target, some form of similarity or structural mapping (e.g., isomorphism, homomorphism) or resemblance between the two is needed. In its most common formulation, the requirement is that an epistemically valuable model must be similar to its target in relevant respects and to a sufficient degree, given its intended use (Giere [1988](#), [1999](#), [2004](#); Mäki [2010](#), [2011](#)). Second, the robustness requirement states that if a model is very sensitive to changes in its seemingly trivial (e.g., tractability) assumptions, its results could be artefacts of these assumptions. Consequently, one cannot confidently carry the lessons learned from a non-robust model to the real world (e.g., Weisberg [2006](#); Kuorikoski et al. [2010](#)).

Such concerns about the use of abstract models in philosophy are reasonable. Nevertheless, judgements concerning representational adequacy and robustness are not easy to make. First, it is not enough to compare a model with its target to evaluate its representational adequacy. Considerations of representational adequacy in the literature cited above focus on pairwise model-target comparisons and isolate the evaluation of a model's epistemic value from its intended uses. However, models are used to serve a diverse set of modelling goals (see e.g., Pielou [1981](#); Wimsatt [1987](#); Odenbaugh [2005](#)), and their representational adequacy can only be judged relative to the intended use. This point is, we hope, uncontroversial. Representational accounts of models frequently explicitly mention the importance of the modelling goals (e.g., Giere [1988](#), [1999](#), [2004](#); Weisberg [2013](#)); nevertheless, the unit of their analysis remains the relation between a model and its target, or the model-target dyad (Knuuttila [2009a](#), [b](#)). That is, representational accounts of models focus on how to provide an account of representational adequacy based on measures such as similarity and resemblance, but provide no guidance on how to take the modelling goals into account.

Second, model-target comparisons are far from straightforward in the case of abstract, theoretically motivated models such as EL models, which depict highly simplified situations that do not correspond to any empirically observable system. Furthermore, there is no simple metric for assessing the "realisticness" of a model's assumptions. Hence, the evaluation of its plausibility is more complicated than is usually assumed.

Third, merely focusing on the similarity between a model and its target also leaves out the context and the model's relation to other models. This goes against the usual scientific practice of considering the added value of a model in the broader context of other—often competing—models, explanations and theories (Ylikoski and Aydinonat 2014). Robustness considerations provide only a partial solution to the problem: robustness analysis only requires focusing on a set of closely related models and overlooks the relevance of the broader range and the context. In sum, merely focusing on pairwise model-target comparisons and robustness leaves out two crucial elements that would help to make sense of the value and limits of models: modelling goals and the context.[2]

If one is to understand the function, value, and limits of a model, one needs to evaluate it relative to its purpose. In the case of philosophical models, the typical goal of the modelling is to support a philosophical argument that is embedded in the context of earlier philosophical arguments and debates.[3] Overlooking this goal severely hinders their proper assessment. Thus, the best way to understand the contribution of models in social epistemology, and in philosophy more generally, is to begin from the argumentative goals of the modellers, and to take into account the argumentative context in their evaluation. From this perspective the model is not, in itself, the argument, it is merely a part of it. In other words, the philosophical argument makes use of the model. To understand such use, one has to reconstruct the argumentative context, which includes not only the intentions and assumptions of the model user, but also the arguments presented in the existing philosophical debate. The latter could also include models introduced in earlier phases of the discussion.

Let us consider Weisberg and Muldoon's model again. They start their paper with an observation about the role and importance of the division of cognitive labour in science:

> While these facts about the nature of contemporary science are well-known to philosophers, having been discussed by Kuhn and Lakatos, among others, surprisingly little has been written about the epistemology of divided cognitive labor and the strategies scientists do and should use in order to divide their labor sensibly (Weisberg and Muldoon 2009, p. 226)

Here they present the broader context of their argument. They also make it clear that the more immediate context includes earlier philosophical models introduced by Kitcher (1990, 1993) and Strevens (2003) as well as the philosophical arguments they support. While expressing their agreement with the general philosophical conclusions derived from these models, Weisberg and Muldoon highlight an unexplored aspect of the division of labour within this context. They develop a model focusing on what

---

[2] We do not claim that all earlier accounts of the use of models in social epistemology ignore modelling goals. For example, Thicke (2019) and Martini and Pinto (2017) explicitly discuss the goals of particular models. However, we introduce a broader conception of the goals and context of modelling and provide an account and a detailed analysis of why and how argumentative goals and context matters.

[3] Here we understand the notion of argument broadly. An argument "is a sequence of statements where all but one of the statements (the premises) are intended to provide evidence, or support, for the remaining statement (the conclusion)" (Cook 2009, p. 15). Philosophical arguments are thus arguments that present conclusions concerning a philosophical topic. Note that this definition allows for both deductive and non-deductive arguments.

happens when scientists in a population adopt different research strategies. Although they acknowledge the highly idealised nature of their model and the lack of robustness checks (Weisberg and Muldoon 2009, p. 250), they nevertheless argue that it is possible to "draw some tentative conclusions about division of cognitive labor" (Weisberg and Muldoon 2009, p. 250). They also claim, on the basis of some additional assumptions as we will show:

> A polymorphic population of research strategies thus seems to be the optimal way to divide cognitive labor (Weisberg and Muldoon 2009, p. 251)

This is the conclusion of their model-based philosophical argument, albeit a tentative one. In sum, Weisberg and Muldoon use their model to contribute to a debate concerning the division of cognitive labour (the argumentative context). Understanding how the model functions as part of their philosophical argument is the key to understanding its value and limits. A model may have various roles in an argument. Among other things, models serve as representations of empirical targets, proof of conceptual possibility, formalised thought experiments and mere illustrations. Thus, the representational adequacy or robustness of a model cannot be assessed independently of the role it is supposed to play in an argument. For example, if the purpose is to support a claim about a conceptual possibility or to explore possibilities, a highly idealized model might suffice. If the purpose is to illustrate the sensitivity of a certain result to certain assumptions, a more complex, non-robust model could do the job.[4] Seeing models as argumentative devices helps in putting their functions into context. What a model does is not its intrinsic feature: it depends on how it is used and what it is used for. Is it used to establish some new claim in the debate, or is the purpose to strengthen or weaken an existing claim? Is it used to examine the generality of a claim, or to demonstrate the irrelevance of certain considerations for the claim? Identifying the *argumentative goal* and the *argumentative modelling moves* (see Sect. 4) that support it helps the reader to determine the intended contribution of a model.

The argumentative context is also relevant in establishing the model's actual contribution, which might turn out to be different from what was intended. The identification of the original intentions of the modeller is of course valuable in assessing whether the model succeeds in satisfying the modeller's goals. However, it is rarely sufficient in understanding the value of a model. An understanding of what the model is good for tends to emerge over time, after careful consideration by other participants in the debate. Thus, the original model, or its presentation alone, could be misleading as a source for estimating its ultimate value (or lack of it). A mere focus on the modeller's intentions overlooks how the model is used in argumentation, which is critical to seeing the flaws in model-based arguments. The argumentative context is decisive, because it helps to shed light not only on what a philosopher wants to do with a model,

---

[4] As an example of highly idealized modelling in social epistemology, consider the opinion-dynamics model by Hegselman and Krause (2009). In order to focus on general features of truth seeking and opinion aggregation—not on any particular target system—the authors adopt what they call a *low-resolution modelling approach* that abstracts away from detail, and only retains a few key properties of judgment formation and deliberative social exchange. As an example of how the use of a more complex model could help philosophers make their point see Frey and Šešelja (2019), who use their model to argue that simplicity of agent-based models may have a price.

but also on what the model *actually* contributes to the debate. This is one reason why philosophers who have expressed their concerns about the Weisberg–Muldoon model and the use of models in social epistemology in general have sometimes missed the point: they have *merely* focused on the properties of a given model instead of considering more carefully how a family of related models could contribute to a broader inquiry.

## 3 EL models in the service of social epistemology

The success of a philosophical argument depends on the strength of the link between the premises and the conclusion. As we stated above, in a model-based argument the premises will cite at least one model. We do not specify *how* the model is used to support an argument, or how it is used to formulate the premises, because this depends on how the philosopher chooses to use it. In the case of EL models, the model used is a *computational* one: a computer program that implements a *conceptual model,* which, in turn, is used to support an *argument*. To understand how EL models function as argumentative devices it is useful to clearly disentangle between these three different layers of the process. As we will show, both modelling layers introduce potential "soft spots" in the argumentation that might be difficult to identify. We will also show where such soft spots reside in EL models, and how they should be taken into account when the strength of the overall argument is assessed.

In this section we first introduce the different layers of argumentation, then we show how they can be used in appraisals of the argument, and in conclusion we discuss an important function of models in philosophical argumentation: establishing difference makers.

### 3.1 The structure of Weisberg and Muldoon's argument

Philosophers employ several kinds of devices in support of their arguments, such as formal derivations, thought experiments, case studies and even common-sense reflection. Models could be considered one of the elements in their toolbox. Both mathematical models (e.g., Kitcher 1990, 1993; Strevens 2003) and computer simulations (e.g., Weisberg and Muldoon 2009; Pöyhönen 2017; Zollman 2007) have been used as such argumentative devices.

As we noted above, in the case of computer simulations, two distinct layers of modelling should be distinguished: the conceptual model and the computational model. A conceptual model is typically presented in a research paper, although some of its details might remain implicit. Weisberg and Muldoon's EL model suggests how the central notions in a general argument, such as *cognitive diversity* and the *division of labour,* can be made precise. The constructs of the *research approach* and the different *learning strategies* of agents, as well as the measures of collective epistemic performance (e.g., *epistemic progress*) are examples of conceptual-model operationalizations of the general notions of interest.
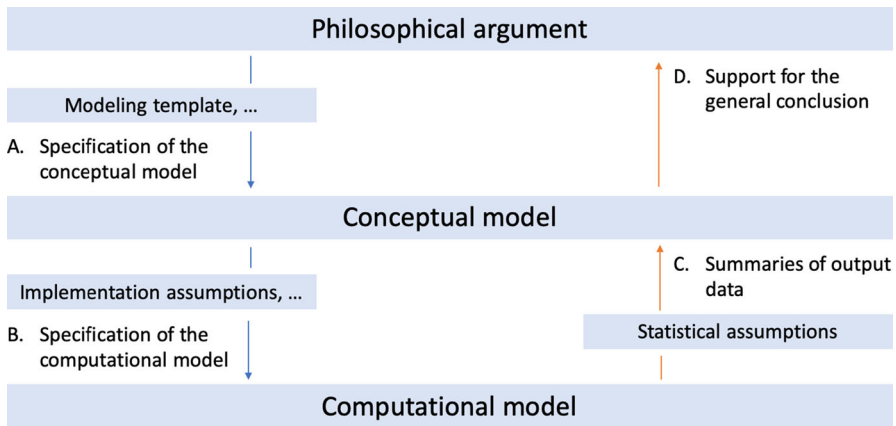
**Fig. 1** Argument and the layers of modelling

In addition, the conceptual model serves as a blueprint for the computational model, meaning the implementation of the model in a computer program. In principle, the computational model could be considered a deductive device that takes the modelling assumptions as input and produces the modelling results as output (Beisbart 2012). It should be borne in mind, however, that even if the computational model is conceived of or reconstructed as a deductively valid argument, this does not imply that the philosophical argument in question is also deductively valid. The model typically supports only part of the reasoning involved in the philosophical argument.

Now let us consider Weisberg and Muldoon's philosophical argument for the following tentative conclusion: the optimal division of labour could be achieved with a polymorphic population of research strategies. How do they support this conclusion? They do so by means of results obtained from the computational implementation of a conceptual model, and two informal assumptions (i.e., assumptions that were not explicitly modelled). Let us now trace how Weisberg and Muldoon arrive at this conclusion. In order to simplify the picture, we omit many details in our skeletal representation of their argument.

In setting up their conceptual model (arrow A in Fig. 1), Weisberg and Muldoon appear to make the following assumptions, among others.[5] For the sake of brevity, we first present the assumption of the conceptual model and then the computational-model analogue of it in parenthesis.

**P1** A scientific research topic can be represented as a set of research approaches (the set of patches comprising the landscape)

**P2** The epistemic significance of research depends on the research approach adopted by the scientist (elevation of the patch)

**P3** Similar research approaches have comparable epistemic significance (smoothness of the landscape)

---

[5] Weisberg and Muldoon present some of these assumptions as the "*conceptual component*[s] of epistemic landscapes" (2009, p. 228, emphasis added).

**P4** Scientists care about epistemic significance, i.e., about the "significance of the truth that is uncovered by employing a given approach" (p. 229) (Agents try to find high-elevation patches on the landscape)

**P5** Scientists can only gradually change their research approach (Agents cannot jump over the landscape: they move one patch at a time)

**P6** Different strategies for changing one's research approach constitute a relevant form of cognitive diversity in a community of scientists (Three types of agents—-controls, followers and mavericks)

…

There is obviously continuity between the conceptual assumptions and the assumptions implemented in the computational model. Nevertheless, the differences are significant. For example, the assumption that scientists care about epistemic significance (conceptual model) differs significantly from the assumption that agents try to find high-elevation patches on a landscape (computational model; see P4, above). Thus, whether the result of the computational model supports the conclusions at the level of the conceptual model turns out to be an important question for the appraisal of the philosophical argument.

Although the assumptions of the conceptual model lay down the basic framework of the analysis, it is still too vague to derive precise results. The implementation of the conceptual model in a computer program (arrow B in Fig. 1) makes the derivation possible, but with a slightly different set of assumptions that are taken to be an appropriate implementation of those of the conceptual model. Moreover, the conceptual model is incomplete in several ways, and it must be accompanied by various implementation assumptions (IA) to make the derivation of quantitative modelling results possible. The original EL model, includes such assumptions concerning:

**IA1** The size of the landscape ($101 \times 101$ grid)

**IA2** Scheduling (do agents move simultaneously or sequentially)

**IA3** Time scale, or how long a simulation is allowed to run (max 50,000 cycles)

…

The addition of these implementation assumptions allows the research approaches, the significance of the research and the behaviour of scientists to be represented in a particular way in the computational model in comparison to the conceptual model presented by Weisberg and Muldoon. It is on these premises and assumptions as well as the corresponding analysis of the computational model that Weisberg and Muldoon (among others) base the following *interim* conclusions (IC, arrow C in Fig. 1):

**IC1** Maverick agents stimulate the problem-solving ability of followers (mavericks help followers to hill-climb)

**IC2** Adding maverick agents to a population of controls and followers increases the population's capacity to make epistemic progress (epistemic progress $=_{df}$ proportion of non-zero patches discovered by the population at a particular time)

**IC3** The increased epistemic efficiency of the mixed population of agents (IC2) relies on mavericks *stimulating* the followers to make considerable epistemic and total progress (IC1)

Note that these interim conclusions still do not imply the tentative conclusion in Weisberg and Muldoon's argument (i.e., that an optimal division of labour could be achieved with a polymorphic population of research strategies). In an attempt to reach this more general conclusion they make an additional informal assumption (IN1), which was neither implemented in the computational model nor presented with the conceptual model.

**IN1** Different strategies have differential costs. In particular, it is more costly to be a maverick than a follower. (p. 250)

On the basis of their interim conclusions and this informal assumption, they conclude (arrow (d)):

**C1** A polymorphic population of research strategies thus seems to be the optimal way to divide cognitive labor. (p. 251)

Furthermore, although not stated in these precise terms, the interim conclusions are meant to support the more general philosophical conclusion:

**C2** Scientific research requires the beneficial division of cognitive labour, which can be brought about by means of cognitive diversity

As this schematic representation of the argumentative structure indicates, the derivation of philosophical conclusions relies on several elements including the specification of models (arrows A–B in Fig. 1) and interpretations of results (arrows C–D in Fig. 1). Each step typically introduces new assumptions into the process. Hence, the strength of the philosophical argument, as well as the truth and relevance of its conclusions, could be contested by challenging the assumptions and the argumentative links, as depicted in our schematic presentation.

### 3.2 Reappraisal of the argument

To see if the string of argumentation works, one could start by asking whether the models are valid in the sense that their results follow from their assumptions. A minimal validity condition for a computational model is that it must be free from programming and implementation errors. The process of ascertaining this is referred to as *verification* in the literature on simulation (see Gräbner 2018 and the citations therein).

It is also worth pointing out that in the case of EL models, what are typically called modelling results are not the same as output data: they are rather summaries of some distribution in the output dataset produced by *numerous runs* of the simulation model. Each single run of a computational model is an algorithmic process, the outputs of which could be considered deductive consequences of its inputs. However, because the modelling results tend to be based on a statistical analysis of numerous runs, they are not simply logical consequences of the modelling assumptions alone. In other words, the appropriate analysis of the output data, the statistical assumptions and

the model used become an important part of the argument. Needless to say, if the statistical modelling of the output dataset is done incorrectly, what are reported as modelling results might not follow from the model's assumptions. As an example, let us consider Weisberg and Muldoon's claim that the epistemic efficiency of the mixed population of agents in their model is due to mavericks *stimulating* the followers to make considerable epistemic and total progress (IC3, above). A careful analysis of the model (see Alexander et al. 2015, Figure 8) reveals that no such stimulation occurs. Instead, further progress on the population level is generated solely by mavericks. The problem is that Weisberg and Muldoon jump to the wrong conclusion due to an insufficiently detailed analysis of the output data.

Another question concerning the strength of the philosophical argument is whether the computational model aligns well with the conceptual model (arrows B and C in Fig. 1). Does the implementation of the simulation line up with the conceptual model? Are the model results robust to changes in the implementation assumptions, parameters, software engineering and decisions, for example? In the Weisberg–Muldoon case, the computational model plays a key role in establishing the mapping from the proportion of mavericks to the dynamics of epistemic progress. Computational implementations of the components of the conceptual model (the different kinds of agents, landscapes, epistemically significant outcomes and so on) are generated in the computer program. Next, the computational model is run with a model structure and parameter values corresponding to different scenarios formulated in terms of the conceptual model. The values of the outcome variables (e.g., epistemic progress) are observed across these runs. Conclusions from the modelling stated on the level of the conceptual model rely on the input–output mapping of the computational model: it is with the help of a computer program that the modeler establishes which conceptual-model conclusions follow from which assumptions. Hence, claims concerning the implications of the conceptual model depend on the computer program and its ability to serve the inferential role attributed to it. In the case of the Weisberg–Muldoon model there is, in fact, a discrepancy between the description of the follower rule in the conceptual model and its computational implementation, which compromises conclusions about the relative epistemic performance of different kinds of agents (Alexander et al. 2015; Pöyhönen 2017). The argument from the premises of the conceptual model to its conclusions is not valid because a sequence of inferential steps implemented in the computational model fails.[6]

Let us now look at the next argumentative step (arrow D in Fig. 1), moving from the conceptual model to philosophical conclusions of interest. To establish that the modelling results alone are not sufficient for making the inferential transition, let us consider, for example, the different conceptual-model constructs used by the authors to capture the idea of epistemic efficiency: they use three distinct outcome measures to track the epistemic efficiency of a population of scientists. First, they keep track of how long it takes the population to find the two *peaks of the landscape*. Second, they define a measure they call epistemic progress as the proportion of non-zero patches visited at a particular time. Finally, they also keep track of the proportion

---

[6] Despite the implementation error, the computational model naturally supports *some* argumentative transition, but it is not the one Weisberg and Muldoon report in their paper, i.e., in their conceptual model.

of all patches visited by some member population, which they label total progress. By providing counterexamples, Pöyhönen (2017) argues that none of these measures can function as an appropriate measure of epistemic efficiency, which is the more general philosophical idea implemented by the three variables. On the one hand, peaks-reached can be maximised by placing all agents in the peak patches, which corresponds to extremely poor epistemic coordination. Epistemic progress, on the other hand, is not sensitive to the elevation of a patch, and hence, especially on rugged landscapes, may correlate poorly with whether the two hills of epistemic significance have been discovered by the population. Finally, what Weisberg and Muldoon call "total progress" appears largely irrelevant to epistemic efficiency because it fails to distinguish the discovery of significant patches from the examination of zero-value approaches.

Note how this problem differs from the ones discussed above. It is an issue that concerns the relationship between the conceptual model and the philosophical argument (arrows A and D in Fig. 1). Even if the computational implementation of the constructs (e.g., epistemic progress) is adequate, the constructs themselves cannot serve the argumentative role assigned to them in the argumentation (i.e., to support conclusions C1 and C2).

Another factor that could weaken the philosophical argument is the lack of robustness of modelling results. If the simulation is highly sensitive to changes in a parameter value or to seemingly insignificant choices in the computational implementation, this challenges the reliability of the general conclusions that could be drawn from it. In this sense concerns about the robustness of philosophical models (see Sect. 2) are justified. However, it should also be recognised that various aspects of a model's robustness emerge over time as other philosopher modellers take on the task of analysing and using the original model. Moreover, the discovery of non-robust results might indicate the conditions under which a philosophical argument is tenable and help in discovering critical assumptions as well as difference makers. Thus, non-robust results could be informative in showing the range of conclusions that could be supported by the computational and conceptual model. As we argue later on (Sects. 3.2 and 5), making sense of the positive and negative implications of non-robustness requires taking the broader argumentative context into account.

Finally, the applicability of the epistemic-landscape conceptual model could also be challenged more generally. The extent to which the philosophical argument is convincing depends on the credibility of the models involved. Appraisal of the credibility of a model requires more than a comparison of the model with its intended target. For example, Sugden (2000) suggests that credibility depends on the logical coherence (i.e., validity) of the model as well on how well the model fits into our general understanding of the causal structure of the world. Credible models, according to Sugden, describe "how the world could be" rather than being accurate descriptions of particular or generalized real-world target systems.[7] Concerning credibility, one could ask the following questions. Are the assumptions of the model plausible given the modelling goals and the argument of which it is a part? Are the construct (model world) and its

---

[7] On the credibility of models also see Cartwright (2009), de Donato Rodríguez and Zamora Bonilla (2009), Grüne-Yanoff (2009), Kuorikoski and Lehtinen (2009), Mäki (2009) and Sugden (2009).

dynamics (what happens in the model world), as well as the model results (findings), broadly consistent with our general knowledge of the world? Obviously, these questions are similar to those concerning representational adequacy discussed above. We are not suggesting that representational adequacy has no role to play in assessments concerning the use of models in philosophy. As should be obvious by now, we are arguing that representational adequacy must be evaluated in light of the argumentative goals and the broader argumentative context. Moreover, one can establish or question the credibility of a model without engaging in pairwise model-target comparisons, by way of making piecemeal assessments of the plausibility of individual assumptions relative to the argumentative goals, and considering how the model and the suggested conclusions fit what we already know about the world. As we will show, in the case of EL models, attempts to improve plausibility and credibility turn out to be another way of exploring what could be called the *argumentative landscape* relevant to the philosophical topic in question, i.e. the constellation of possible premises, conclusions and paths of reasoning connecting them. Such exploration results in a better understanding of the dependency between proposed variables, and of the conclusions that a given set of premises can support. However, to show this we should first discuss an important function of modelling: establishing difference makers.

### 3.3 Difference making and robustness

We suggested above that understanding how models function as argumentative devices requires making explicit the assumptions that contribute to model-supported arguments and keeping score of how strongly they support the conclusion of interest. Grasping the different steps of the argumentation and the credibility of the models is only part of the story, however. Bringing the value-added of simulation models into focus requires closer attention to the kind of conclusion being argued for. Simulation-based philosophical arguments typically use models to support claims about difference-making. For example, Weisberg and Muldoon aimed to show in their modelling efforts that *changing* the proportion of mavericks in the population of scientists *makes a difference* to its epistemic performance. Later models in the EL tradition qualified this difference-making result in various ways. Thoma (2015), for example, shows that such dependency holds only when scientists are not too inflexible in their choice of a new research topic and not too ignorant of others' work. Pöyhönen (2017) suggests that the original difference-making relation between cognitive diversity and epistemic performance does not hold on the kind of smooth landscapes studied by Weisberg and Muldoon; it is only on rugged landscapes that a heterogenous population of agents outperforms a homogenous population. Although pointing in somewhat different directions, all such results concern a difference-making relationship between (changes in) a modelling assumption, and (changes in) an outcome of interest.

Establishing difference making goes beyond the task of establishing validity. Showing that a modelling assumption makes a difference to an outcome relies on the various ways of introducing variation into the assumptions. Analogously with Mill's method of difference (Mill 1974, Book III, §2), to show that something makes a difference to something, one must find a pair of scenarios that are similar in all aspects except one,

such that the difference in the input variable leads to a change in the output. EL modelling includes exactly this kind of reasoning. A model is run with different values of the independent variables of interest, and the effect of this variation on the dependent variable of interest is observed. A minimal condition for establishing such a difference-making claim is that holding other modelling assumptions constant, changes in the value of the independent variable should lead to changes in the dependent variable.[8] A simple example of this is mentioned above: holding other things constant, Weisberg and Muldoon ran their model with different proportions of mavericks, and observed the resulting changes in epistemic progress (Weisberg and Muldoon 2009).

Nevertheless, such minimal difference-making tends to be of little argumentative value because it can only support the claim of dependence under a very specific set of modelling assumptions. For most purposes, the relationship between two variables should have some generality across the variation in the variables, and across different values of other parameters of the model. As critics have pointed out, it was difficult to determine the relevance of the results achieved by Weisberg and Muldoon because the analysis of the original EL model did not include any sensitivity or robustness analysis. In other words, even if the proportion of mavericks makes a difference to epistemic progress under the set of parameter values employed by Weisberg and Muldoon, it may be that such an effect only occurs in this particular part of the parameter space of the model, and hence the difference-making relation might be highly local, even pointlike. Such locality (non-generality) makes it questionable whether a model can be used to support more general conclusions.[9] Establishing the generality and derivational robustness of a difference-making result requires showing that some parts of the model are *not* difference makers as far as the result is concerned. For example, and intuitively, results from EL models should not be sensitive to the size of the landscape, the exact position, shape and number of the hills of epistemic significance, or the order in which agents move during each time step.

Establishing the irrelevance of such factors works differently in analytical and simulation models. In the case of analytical modelling, similar results are often derived under less stringent assumptions during the evolution of a modelling framework, leading to more general results. In agent-based modelling, however, it is typically not possible to eliminate modelling assumptions so as to make the results more general. All the details of the model must be implemented in some form in order to make the computational model run, thus the generality of results must be established by some other means. In the following section we suggest that sensitivity analysis, robustness analysis and the construction of model families could be considered argumentative modelling moves to that end. More generally, understanding the argumentative role of agent-based models in philosophy requires not only an understanding of model

---

[8] Note that in ABMs or in complexity modelling, changes in modelling assumptions are often not independent/additive. Changing the value of one parameter often influences how other parameters change outcomes, which means that their relation is nonlinear. This implies that for most purposes, one ought not simply change these variables one at a time. Instead, the "web of dependencies" must be charted by introducing variation that changes several things at once.

[9] Rosenstock et al. (2017) raise a similar concern about Zollman's (2007) results showing that the structure of the epistemic network is connected to the premature convergence of opinions only when the payoffs of the two hypotheses are extremely close to one another ($p_A = 0.5$, $p_B = 0.501$ in the original model).

structure but also practices of model construction and use. The epistemic benefit to be derived from modelling is commonly realised as a consequence of several modelling attempts—rather than from the development of a single model (more on this in Sect. 5).

## 4 Modelling moves as argumentative moves

We now introduce the notion of argumentative move and discuss how changes to a model made by later authors can be analysed as such. This will shed light on the process of collective exploration, which ultimately determines the argumentative value of a model. We will also show how changes in a model may give rise to second-generation models that are intended to support philosophical conclusions quite different from the original ones.

We define *argumentative goals* as the general and interim philosophical conclusions that a philosopher wishes to establish. Such goals are not always explicitly stated, and they may be presented partially or in a vague manner. It could be that the aims of the argumentation are so clear from the context that there is no need to articulate them explicitly. In any case, for the purposes of analysis it is useful to try to articulate such argumentative goals. We do not presume that a philosopher first chooses the argumentative goal and then proceeds to the details of the argument. He or she may, for example, explore a model and discover a goal that can be furthered by using, modifying and applying it. In this paper, we use the notion of an argumentative goal to reconstruct the product, not to hypothesise about the process by which it was reached.

*Argumentative moves* are what philosophers do to reach argumentative goals, given the context of existing arguments. Such moves include criticism of and amendments to existing arguments, as well as the introduction of new ones.[10] They may, for example, seek to:

(a) demonstrate the possibility or impossibility of something;
(b) introduce a new idea or consideration into the debate;
(c) examine and hence establish or challenge the validity, generality or scope of an earlier argument;
(d) support or undermine earlier claims about difference-making;
(e) modify an earlier argument to correct mistakes or to make it more plausible;
(f) provide additional arguments supporting either the premises or the conclusions of an earlier argument;
(g) broaden the debate by introducing a new perspective on the problem in question.

We do not claim that this list is exhaustive or that the moves are exclusive alternatives. The list could be extended quite easily. Given that our interest lies in how models are used in philosophical argumentation, we focus on the argumentative moves that involve models. We refer to these as *modelling moves*.

---

[10] Making some of the argumentative moves can be conceived of as interim argumentative goals. Nevertheless, we wish to keep these separate to emphasise what the philosophers *do* to establish and support their arguments.

Modelling moves include (but are not limited to):

(1) modifying the assumptions of the conceptual model;
(2) implementing the assumptions of a conceptual model in a computational model;
(3) articulating the informal assumptions leading to the more general philosophical conclusion;
(4) introducing a novel set of assumptions and a new model.

We have discussed several modelling moves that concern the Weisberg–Muldoon model and have observed, for example, how philosophers question the various steps in Weisberg and Muldoon's argumentation by modifying the original model and demonstrating its problems. A significant set of argumentative moves in the literature that followed the publication of the EL model contains moves that challenged the validity, generality and scope of their argument. The related modelling moves included modifying the assumptions and the parameter values in the computational model to examine its robustness and sensitivity. Robustness and sensitivity analyses could be seen as tools that help one to find one's way around an existing web of arguments. Finding highly robust results, in other words conclusions that can be sustained on a wide variety of premises, is of course valuable. However, such results are rare. What the modeller often discovers is that a certain result is not robust to a change in premises: in other words, a new conclusion follows from a modified set of premises. Nonetheless, discovering a lack of robustness is valuable because it helps modellers learn about the argumentative landscape, and discover what conclusions follow from specific sets of premises.

Although Weisberg and Muldoon's original paper did not report on the robustness of their results, the work of Thoma (2015), Alexander et al. (2015) and Pöyhönen (2017) all contribute to such an evaluation. For example, running versions of the model on both three-dimensional rugged landscapes and NK landscapes provides robustness checks on Weisberg and Muldoon's original claims about the usefulness of diversity. Alexander and co-authors, using the new NK model, show that cognitive diversity is not necessarily beneficial to an epistemic community, and that it could also do harm. Pöyhönen's results similarly qualify Weisberg and Muldoon's original claim: in an attempt to make it more precise, the author suggests that cognitive diversity only provides epistemic benefits on rugged landscapes, not on smooth ones.[11]

The introduction of a modified or a new model does not only concern robustness analysis, however, and commonly serves to increase the plausibility of the assumptions involved in the argument chain. In addition to raising concerns about the generality of the modelling results, such modelling moves could also be seen as instances of

---

[11] In simulation modelling, examining the validity of models also requires dealing with the problem of inferential opacity (Humphreys 2004). Given that it is not possible to inspect all the algorithmic steps in a model run, the modeller must resort to other techniques in building a convincing case that modelling results do, in fact, follow from the stated assumptions. Using well-known submodels, unit testing, and running simpler versions of the model are ways of dealing with the credibility challenge in a piecemeal fashion, arguing for the reliability of the model by establishing the reliability of individual components. Rigorous and transparent data analysis on multiple levels is another way of addressing the transparency challenge. As mentioned in Sect. 3, it was through careful data analysis that Alexander and co-authors challenged Weisberg and Muldoon's claim about the mechanism by which diversity increases epistemic performance. As they showed, the existence of mavericks does not stimulate follower performance, contrary to Weisberg and Muldoon's assertion.

challenging the applicability of the model or showing the limited scope of the argument. Such challenges typically concern the relevance of the modelling assumptions or the interpretation of the results. Thoma, for example, challenges WM's original model by arguing against premise P5 (see Sect. 3.1) according to which scientist agents can only move within their Moore neighbourhood (2015, p. 462), claiming that this implies an "extreme level of short sightedness and inflexibility among scientists" (2015, p. 462). She shows that alternative search rules for agents, which appear to be just as compatible with the behaviour of scientists as those suggested by Weisberg and Muldoon, lead to clearly different collective outcomes. Having made these moves, she feels able to suggest that a division of labour is only beneficial when scientists are not too inflexible in their choice of a new research topic and not too ignorant of other people's work. In sum, Thoma not only shows the limitations of an existing argument, but also develops a model that supports a more intuitive argument:

> The model I have presented not only supports an intuitive result that Weisberg and Muldoon's could not; *it is also more credible* than Weisberg and Muldoon's in two major ways: First, it is not restricted to local movement, which I have argued is implausible as a representation of scientific practice. Second, the explorer and extractor strategies are better descriptions of the behavior of scientists than the maverick and follower strategies, since both explorers and extractors avoid the mere duplication of work others have done. What further speaks in favor of the model is that a number of its implications map ccredibly [sic] onto features of actual scientific practice, as evidenced in the course of the article. For instance, on my model it turns out to be explorer-type behavior that needs special incentives, which seems plausible. (Thoma 2015, p. 471, emphasis added)

Alexander et al., on the other hand, show the limits of Weisberg and Muldoon's model and hence the limited applicability of their conclusion by employing a more plausible assumption:

> The crucial difference between the NK model just described and Weisberg and Muldoon's epistemic landscape is this. Expressed as an NK model, Weisberg and Muldoon's model assumes fitness functions are highly correlated. [...] Why does this matter? It matters because the Weisberg and Muldoon model builds into the basic topology of the epistemic landscape correlations that make social learning advantageous. As such, we should not be surprised to find, in the case they consider, that cognitive diversity and social interactions between agents can be beneficial. But, as the generalization to NK landscapes shows, social learning is not always beneficial. Whether social learning is beneficial or harmful depends on the topology of the epistemic landscape, a point of which we know very little. (Alexander et al. 2015, p. 448, emphasis added)

Hence, argumentative moves are not limited to modelling moves that purport to show the flaws in an existent argument. One might also amend the model to make the argument more precise, or add new features in arguing for a different point. Alexander et al.'s model is a good example of a positive contribution that, at first sight, appears merely as a critical examination. The authors go on in their paper to derive estimates of the upper and lower bounds of sensible search time on the landscape. In addition

to providing a starting point for their critical examination of the original EL model, the methods they employ also give useful insights into the general properties of EL model worlds.

Similarly, the broadcasting model developed by Pöyhönen (2017) extends the EL framework in novel ways. The landscapes studied are dynamic, in that the epistemic work done by the agents alters the distribution of epistemic significance "mass" between the different research approaches. Consequently, it is implied that cognitive diversity may lead to collective epistemic benefits by bringing about beneficial coordination between members of the epistemic community (i.e. by modulating the exploration–exploitation trade-off in collective search).

The literature published after Weisberg and Muldoon introduced their model shows that, despite the shortcomings of the original model, the idea of interpreting a research topic as a landscape and modelling scientists as agents with different strategies has given rise to a new and fruitful framework, and has raised a number of research questions that had not thus far been carefully analysed. A further indication of the fruitfulness of the approach is that a "second generation" of models has emerged since the publication of the first set of models building on the original EL framework (Alexander et al. 2015; Thoma 2015; Pöyhönen 2017). In many of these second-generation models, the EL framework is applied beyond its original domain to study issues such as funding allocation in science and to derive science-policy-relevant conclusions. The models also provide more examples of argumentative moves.[12]

Whereas the first-generation models focused on refining and criticising the original Weisberg–Muldoon model, the second-generation models are modified versions of the original model to serve different argumentative goals. For this reason, it is important to pay close attention to their details and argumentative goals. For example, Shahar Avin's (2019) main aim is to examine the role of funding-allocation mechanisms. Deviating from the original goals of the Weisberg–Muldoon model, Avin primarily attempts to direct the attention of social epistemologists from individual motives and learning strategies to institutional arrangements. At the same time, he addresses a novel audience, science-policy researchers, to convince them that the employment of lottery mechanisms in funding allocation could lead to a better epistemic output. Despite the introduction of new argumentative goals, Alvin's argument is based on relatively modest modifications to the Weisberg–Muldoon model. This is crucial: if one were to focus merely on the features of Alvin's model without paying close attention to the argumentative goals and context, one would easily miss the point of the exercise.

The second-generation models developed by Balietti et al. (2015) and Currie and Avin (2019) provide another kind of example, where the models as well as the argumentative goals and the conceptual premises are quite different. Just as Weisberg and Muldoon borrowed their argumentative device (the idea of an adaptive landscape and the techniques to model it) from biology and adapted it to serve their own purposes (Gerrits and Marks 2015), these authors transform the EL framework to suit their purposes. As in the case of Balietti et al., the audiences can be remarkably different too. Their article was published in PLOS ONE, not a typical forum for social

---

[12] The boundary between first- and second generation models is not always clear-cut. Harnagel (2019) is a good example, and could be interpreted as either, depending on which aspects of the paper are emphasised.

epistemology. These changes make the arguments of the second-generation models largely independent of Weisberg and Muldoon's argument. Whatever is the ultimate scholarly evaluation of Weisberg and Muldoon's argument, it does not directly affect, strengthen or weaken the argumentative force of the second-generation models. In other words, their epistemic fates are different. As a consequence, the analysis of the implications of the second-generation models could be carried out separately from the original discussion. Nevertheless, second-generation models might still be useful for philosophers. For example, modellers in philosophy could learn a lot from Balietti et al.'s methodology and carefully executed analysis, even though they might have difficulties in accepting their implementation of the idea of "ground truth."

The second-generation models appear to expand the space of exploration. However, given the combination of modified EL models and different argumentative goals, they also introduce new ambiguity to the interpretation of EL models. From the perspective of the modellers, one way to overcome this ambiguity is to make the argumentative steps associated with agent-based models more explicit than is currently the practice. Using models as black-boxes in philosophical argumentation and leaving many steps of the argumentation implicit often result in barriers to understanding. If the changes to the modelling assumptions, the argumentative goals and the way in which the model is used remain implicit, the nature of the modelling endeavor might remain obscure and lead to misunderstandings. Moreover, the lack of explicitness could invite skepticism as the critical audience often takes the lack of argumentative detail as an indicator of arrogance and obscurity. Finally, as new models and arguments are introduced, the contribution of the first-generation models may be blurred as the (possibly unrecognised) second-generation models muddle the debate. As it is in the case of science, models in philosophy should not be used as black boxes in argumentation. It is crucial that the audience understands what goes inside the model (the relation between the conceptual and the computational), how the model is used to reach the conclusions, and how this fits to the broader argumentative context.

## 5 The model family as a unit of epistemic evaluation

Thus far, the discussion has focused on particular argumentative goals and modelling moves that are to be found in articles published in the wake of Weisberg and Muldoon's original paper. This helped us to highlight the role of models as argumentative devices. However, the narrow focus on individual modelling moves still keeps us from fully grasping the value and limits of models in social epistemology. With this focus, each argumentative move appears as a countermove to a previous one, whereas the ultimate raison d'être behind the use of the EL framework in social epistemology is an interest in cognitive diversity. To avoid being blindsided by individual argumentative modelling moves, one could zoom out from individual models and articles to see how the argumentative moves and the individual models relate to one another, and consider whether they—as a whole—serve any useful purpose in terms of understanding the role of cognitive diversity in science.

Table 1 summarises the assumptions and results of the first-generation EL models. One way of reading the table is to look at the individual models and their results.

**Table 1** Assumptions and results of EL models

|  | (New) Assumptions | Results |
|---|---|---|
| Weisberg and Muldoon (2009) | Control, follower and maverick search rules<br>Smooth landscapes<br>Measures of epistemic efficiency: (1) finding peaks, (2) epistemic progress and (3) total progress | Division of labour: epistemic performance of a population increased when maverick agents are added to it<br>Pure population of maverick agents outperforms all other populations |
| Thoma (2015) | Local search is not constrained to the Moore neighbourhood<br>Extractor and extractor search rules | Division of labour is only beneficial when scientists are not too inflexible in their choice of a new research topic and not too ignorant of others' work |
| Alexander et al. (2015) | Swarm rule<br>Study of NK landscapes | Diversity can adversely affect epistemic efficiency<br>Swarm agents are more efficient than mavericks<br>Social learning only works on particular NK landscapes |
| Pöyhönen (2017) | Broadcasting rule for social learning<br>Measuring performance as epistemic work<br>Rugged landscapes | No social learning is needed on smooth landscapes<br>Mixed populations fare better than 100% maverick populations on rugged landscapes<br>Diversity modulates a collective exploration–exploitation balance |

Unrealistic and non-robust models follow one another, and it is not clear what their contribution is. Alternatively, one could see the table as a summary of argumentative modelling moves that philosophers employ to make a point in a debate. Although this reading is more useful than the first one—because it pays attention to the argumentative context—it still does not help the reader to fully appreciate the value of the individual models. A third option is to zoom out from the individual models and focus on the table as a whole. Consequently, one can see that there is more to the individual contributions than serving as counterarguments to an existing model. We adopt the third reading and view these models as a family that helps philosophers to explore an argumentative landscape. The modelling efforts of individual philosophers could be seen as attempts to explore and chart a network of dependency relations between modelling assumptions and conclusions of interest.

We argue above that each of these models helps in establishing difference makers within the framework defined by their assumptions. Focusing on individual models, however, prevents us from seeing the epistemic benefit of modelling that is realised as a consequence of several modelling attempts. This is because the individual models do not support a single strong conclusion about cognitive diversity and epistemic performance. In fact, each model shows what happens given its particular set of assumptions.

In the absence of knowledge about what happens when these assumptions change, it is difficult to see if the results of the model have wider applicability. However, conceived of as a family, individual models introduce variation to Weisberg and Muldoon's EL framework. In other words, as a family, EL models help to make visible a set of dependencies that can be established under a variety of related assumptions. The collection of individual argumentative modelling moves within the model family helps in mapping a web of dependencies between changes in subsets of modelling assumptions and their implications.

Let us consider how this differs from the derivational robustness of *a* modelling result. Non-robust results are commonly considered a weakness of a model. However, from the family perspective they could be helpful in terms of articulating the scope of dependencies indicated by individual models, and thereby direct attention to unexplored dependencies in the modelling framework. Thus, what might seem to be a troubling non-robust result in an individual model could provide valuable information when considered in the context of a family of models. Introducing variation into the modelling environment could help modellers to map their way in an argumentative landscape, and to discover which conclusions follow from which sets of premises. In other words, from the family-of-models perspective what matters is the discovery of the variety of dependencies between premises and conclusions. The question at the family-of-models level is not whether the individual model results are derivationally robust, but what general claims one can support on the basis of the available family of models.[13]

Modelling in this sense could be conceived of as learning about model worlds and exploring an argumentative landscape. It is no surprise that modellers themselves often emphasise the explorative nature of a modelling endeavour. Weisberg and Muldoon, for example, are aware of the limitations of their model and suggest that the EL framework needs further exploration. They admit that their model "only scratch[es] the surface of what might be explored using epistemic landscape models" (2009, p. 249), and suggest ways in which to explore this framework:

> Landscapes can be made more rugged, they can contain more information, exploration strategies can take into account more information, an economy of money and credit can be included, and so forth. Much work remains to be done in realizing these possibilities, all of which we believe can be built within our existing framework. (2009, p. 249)

The family-of-models perspective also fosters the realisation that variation in models is not limited to small changes in the respective assumptions. The models differ on several dimensions: central concepts (diversity, division of cognitive labour, social learning) are operationalised differently, different mechanisms function between diversity and epistemic success, and the models keep track of various outcome measures. It is difficult to make sense of this solely from the perspective of derivational robustness, because the set of EL models do not result from systematic derivational robustness analysis, which would require altering one assumption at a time. However, if the vari-

---

[13] We thank one of the referees of the paper for pointing out that at the family-of-models level, a broader notion of robustness (rather than the more stringent derivational robustness) could be employed.

ation between these models is seen as the consequence of argumentative moves that serve argumentative goals, it is easier to understand why they differ on several dimensions. The individual modelling attempts are not only concerned with the robustness of one existing model. As we have pointed out, contributions to the literature also purport to explore what can be argued on the basis of a distinct set of more plausible assumptions, or what happens if new considerations, such as funding, are introduced into the modelling framework. For example, not only do Alexander et al. (2015) explore the Weisberg–Muldoon model to enhance understanding of the dependencies involved and to check whether the model is interpreted appropriately to fit the subject of the study, they also introduce what they think of as a more plausible model to analyse the dependencies.

If one focuses on EL models as a cluster, without losing sight of the argumentative goals served by each one, it is easier to see how the particular argumentative moves help to identify possible dependencies among the selected set of factors. Not only does exploration of the EL framework enhance understanding of how individual models work, it also helps in reasoning about more general dependencies. In fact, building multiple idealised models could be helpful in terms of focusing on a limited set of factors at any one time, and as a strategy it helps to overcome the complexity of the subject matter (Weisberg 2013). In this sense, building a family of models could be conceived of as a collective (although not usually intentional) argumentative move, which helps to establish and refine a philosophical conclusion of interest

Despite its flaws, the Weisberg–Muldoon model provided a new approach to a problem and helped an expanding community of modellers to explore different ways in which cognitive diversity could influence epistemic outcomes. What this amounts to in family-of-models terms is a network of *what-if*s that matches different aspects of the population with different kinds of epistemic success. Even though the EL family does not support a "master" conclusion, it has in an incremental fashion enhanced understanding about the various dependencies between the elements of the modelled scenario.

Note that the family-of-models perspective may lead to an epistemic evaluation of the model's contribution that differs from the one drawn up by the people proposing it. It might turn out that the argumentative moves the modeler first had in mind are more limited in terms of usability than they envisioned, or even unviable, and a collective process of exploration might identify better and more credible applications of the same basic ideas.

As we point out above, the two most common criteria indicating the epistemic value in models, representational adequacy and derivational robustness, ignore the argumentative context and hence do not suffice to shed light on how an abstract model contributes to a philosophical debate, or on how it helps to answer a philosophical question. The family-of-models perspective further supports this argument in implying that one cannot assess the argumentative contribution of a model by focusing merely on its first presentation. The argumentative force of the Weisberg–Muldoon model becomes clear only as the result of the collective work of the people who refined, criticised and explored the EL framework. The understanding gleaned from working with these model variants cannot be summarised in one model. Knowledge of model

variants accumulates gradually, enhancing understanding of the web of dependencies among the assumptions, selected factors and the conclusions.

## 6 Conclusions

We have argued that models in philosophy should be seen as argumentative devices. As the case of EL models demonstrates, they do not necessarily have concretely identified targets and they are not intended to be used as accurate representations. Furthermore, pairwise model-target comparisons alone cannot reveal much about the value of a (set of) models because they omit the relevant argumentative goals ("what the model is used for"), the argumentative context and important details concerning the use of models in a philosophical argument. Analyses of models should include the argumentative context because their full epistemic value can only be perceived when their use is seen as part of argumentative exchange.

We have also argued that the different EL models constitute a cluster or family of models, each change in a model reflecting an argumentative move made by the modeller. Each move serves an argumentative goal: supporting or debunking an existing argument/conclusion, evaluating its scope, or extending an argument pattern into a new domain, for example. Furthermore, this model family should be considered as a whole if the full epistemic contribution of the models is to be understood: the understanding created by a family of models cannot be summarised in any one of them. What the study of model variants contributes is a piece-by-piece accumulation of knowledge about dependencies between assumptions and results. As the family-of-models perspective shows, the general contribution of a model can be better assessed after systematic variations of it have been studied. This is highlighted in the argumentative perspective: the simulation result is not the same as the argumentative conclusion that the simulation model was intended to support.

The perspective outlined in this paper could be fruitful as a method for rationally reconstructing the argumentative contribution of agent-based simulations and other abstract models in philosophy and in science. Rather than getting stuck with the apparent representational inadequacies of these models, or hard-to-interpret authorial intentions, in this approach the attention is directed to issues that really matter in terms of evaluating the model's contribution. As our reconstruction of the debate about EL models indicates, it is not uncommon for the model's true value to be revealed only after a series of replications, modifications and extensions. Although the significance of the original idea cannot be denied, the full epistemic contribution is a collective product.

It is notable that the approach developed in this paper is also applicable beyond social epistemology and philosophy. Assessment of the epistemic contribution of highly simplified theoretical models is also difficult in the sciences.[14] Modellers in the sciences may have different argumentative goals from those of their philosophical cousins, but general ideas about the importance of the argumentative context, the

---

[14] Frigg and Nguyen (2017) provide an overview of the difficulties faced by representational accounts of models.

implicit steps between the model and the intended theoretical conclusions, the significance of argumentative and modelling moves, and the contribution of the collective exploration of model variants apply in both contexts. The utility of a general approach such as this is especially salient when the modelling frameworks travel—as in the case of EL models (Gerrits and Marks 2015)—from one discipline to another.

One advantage of the proposed approach is that it makes it possible to compare simulation models to other argumentative resources employed in philosophy. Thought experiments are an interesting comparison, because they are widely used in epistemology, and models could be considered formalised thought experiments (Currie and Avin 2019). In contrast to thought experiments, model development enables the model's premises to be related to the conclusions in a systematic and rigorous way: unlike in thought experiments, the relation between assumptions and conclusions depends not on reader's intuition, but on explicitly stated (and ideally freely available) premises and implementation detail, and can be exposed to public scrutiny.

Simulation models deserve a fair hearing, and we claim that only by adopting an argumentative approach such as the one outlined in this paper is it possible thoroughly to assess their contribution to social epistemology and philosophy in general. In our judgment, it is too early to say how valuable an argumentative resource this particular class of argumentative devices might be. In any case, the contribution of these models should be assessed accurately. Hyping about their potential could easily lead to the overestimation of their argumentative reach, which could produce a backlash that could undermine their serious use in later debates. For example, to argue that social epistemological toy models suffice to justify science-policy recommendations is quite premature (pace Kummerfeld and Zollman 2015). On the other hand, those who are highly critical of models run the risk of missing out on many valuable contributions. Given that social epistemology, like all areas of philosophy, is seeking to employ strong and credible argumentative resources that produce both *convincing* and *interesting* conclusions, we should not be purists with respect to available argumentative devices.

**Authors' contributions** This work is fully collaborative.

# References

Alexander, J. M., Himmelreich, J., & Thompson, C. (2015). Epistemic landscapes, optimal search, and the division of cognitive labor. *Philosophy of Science, 82*(3), 424–453. https://doi.org/10.1086/681766.

Avin, S. (2019). Centralized funding and epistemic exploration. *The British Journal for the Philosophy of Science, 70,* 629–656. https://doi.org/10.1093/bjps/axx059.

Balietti, S., Mäs, M., & Helbing, D. (2015). On disciplinary fragmentation and scientific progress. *PLoS ONE, 10*(3), e0118747. https://doi.org/10.1371/journal.pone.0118747.

Bedessem, B. (2019). The division of cognitive labor: Two missing dimensions of the debate. *European Journal for Philosophy of Science, 9*(1), 3. https://doi.org/10.1007/s13194-018-0230-8.

Beisbart, C. (2012). How can computer simulations produce new knowledge? *European Journal for Philosophy of Science, 2*(3), 395–434. https://doi.org/10.1007/s13194-012-0049-7.

Cartwright, N. (2009). If no capacities then no credible worlds. But can models reveal capacities? *Erkenntnis, 70*(1), 45–58. https://doi.org/10.1007/s10670-008-9136-8.

Cook, R. T. (2009). *A dictionary of philosophical logic*. Edinburgh: Edinburgh University Press.

Currie, A., & Avin, S. (2019). Method pluralism, method mismatch & method bias. *Philosophers' Imprint, 19*(13), 1–22.

de Donato Rodríguez, X., & Zamora Bonilla, J. (2009). Credibility, idealisation, and model building: An inferential approach. *Erkenntnis, 70*(1), 101–118. https://doi.org/10.1007/s10670-008-9139-5.

Frey, D., & Šešelja, D. (2018). What is the epistemic function of highly idealized agent-based models of scientific inquiry? *Philosophy of the Social Sciences, 48*(4), 407–433. https://doi.org/10.1177/0048393118767085.

Frey, D., & Šešelja, D. (2019). Robustness and idealizations in agent-based models of scientific interaction. *The British Journal for the Philosophy of Science*. https://doi.org/10.1093/bjps/axy039.

Frigg, R., & Nguyen, J. (2017). Models and representation. In L. Magnani & T. Bertolotti (Eds.), *Springer handbook of model-based science* (pp. 49–102). Dordrecht: Springer.

Gerrits, L., & Marks, P. (2015). The evolution of Wright's (1932) adaptive field to contemporary interpretations and uses of fitness landscapes in the social sciences. *Biology and Philosophy, 30*(4), 459–479. https://doi.org/10.1007/s10539-014-9450-2.

Giere, R. N. (1988). *Explaining science: A cognitive approach*. Chicago: University of Chicago Press.

Giere, R. N. (1999). Using models to represent reality. In L. Magnani, N. J. Nersessian, & P. Thagard (Eds.), *Model based reasoning in scientific discovery* (pp. 41–57). New York: Kluwer Academic Publishers.

Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science, 71*(5), 742–752. https://doi.org/10.1086/425063.

Gräbner, C. (2018). How to relate models to reality? An epistemological framework for the validation and verification of computational models. *Journal of Artificial Societies and Social Simulation*, *21*(3), 8. https://doi.org/10.18564/jasss.3772.

Grüne-Yanoff, T. (2009). Learning from minimal models. *Erkenntnis, 70*(1), 81–99. https://doi.org/10.1007/s10670-008-9138-6.

Harnagel, A. (2019). A mid-level approach to modeling scientific communities. *Studies in History and Philosophy of Science Part A, 76,* 49–59. https://doi.org/10.1016/j.shpsa.2018.12.010.

Hegselman, R., & Krause, U. (2009). Deliberative exchange, truth, and cognitive division of labour: A low-resolution modeling approach. *Episteme, 6*(2), 130–144.

Humphreys, P. W. (2004). *Extending ourselves: Computational science, empiricism, and scientific method*. Oxford: Oxford University Press.

Kitcher, P. (1990). The division of cognitive labor. *Journal of Philosophy, 87*(1), 5–22. https://doi.org/10.2307/2026796.

Kitcher, P. (1993). *The advancement of science: Science without legend, objectivity without illusions*. New York: Oxford University Press.

Knuuttila, T. (2009a). Isolating representations versus credible constructions? Economic modelling in theory and practice. *Erkenntnis, 70*(1), 59–80. https://doi.org/10.1007/s10670-008-9137-7.

Knuuttila, T. (2009b). Representation, idealization, and fiction in economics: From the assumptions issue to the epistemology of modeling. In M. Suárez (Ed.), *Fictions in science: Philosophical essays on modeling and idealization* (pp. 205–231). London: Routledge.

Kummerfeld, E., & Zollman, K. J. (2015). Conservatism and the scientific state of nature. *The British Journal for the Philosophy of Science, 67*(4), 1057–1076.

Kuorikoski, J., & Lehtinen, A. (2009). Incredible worlds, credible results. *Erkenntnis, 70*(1), 119–131. https://doi.org/10.1007/s10670-008-9140-z.

Kuorikoski, J., Lehtinen, A., & Marchionni, C. (2010). Economic modelling as robustness analysis. *The British Journal for the Philosophy of Science, 61*(3), 541–567. https://doi.org/10.1093/bjps/axp049.

Mäki, U. (2009). Missing the world. Models as isolations and credible surrogate systems. *Erkenntnis, 70*(1), 29–43. https://doi.org/10.1007/s10670-008-9135-9.

Mäki, U. (2010). Models and truth: The functional decomposition approach. In M. Suárez, M. Dorato, & M. Rédei (Eds.), *EPSA epistemology and methodology of science: Launch of the European Philosophy of Science Association* (pp. 177–187). Berlin: Springer.

Mäki, U. (2011). Models and the locus of their truth. *Synthese, 180*(1), 47–63. https://doi.org/10.1007/s11229-009-9566-0.

Martini, C., & Pinto, M. F. (2017). Modeling the social organization of science: Chasing complexity through simulations. *European Journal for Philosophy of Science, 7*(2), 221–238. https://doi.org/10.1007/s13194-016-0153-1.

Mill, J. S. (1974). *A system of logic. Books I–III*. Toronto: University of Toronto Press.

Odenbaugh, J. (2005). Idealized, inaccurate but successful: A pragmatic approach to evaluating models in theoretical ecology. *Biology and Philosophy, 20*(2–3), 231–255. https://doi.org/10.1007/s10539-004-0478-6.

Pielou, E. C. (1981). The usefulness of ecological models: A stock-taking. *The Quarterly Review of Biology, 56*(1), 17–31.

Pinto, M. F., & Pinto, D. F. (2018). Epistemic landscapes reloaded: An examination of agent-based models in social epistemology. *Historical Social Research, 43*(1), 48–71. https://doi.org/10.12759/hsr.43.2018.1.48-71c.

Pöyhönen, S. (2017). Value of cognitive diversity in science. *Synthese, 194*, 4519–4540. https://doi.org/10.1007/s11229-016-1147-4.

Reijula, S., & Kuorikoski, J. (2019). Modeling epistemic communities. In M. Fricker, P. J. Graham, D. Henderson, N. Pedersen, & J. Wyatt (Eds.), *The Routledge handbook of social epistemology* (pp. 240–249).

Rosenstock, S., Bruner, J., & O'Connor, C. (2017). In epistemic networks, Is less really more? *Philosophy of Science, 84*(2), 234–252. https://doi.org/10.1086/690717.

Strevens, M. (2003). The role of the priority rule in science. *Journal of Philosophy, 100*(2), 55–79. https://doi.org/10.5840/jphil2003100224.

Sugden, R. (2000). Credible worlds: The status of theoretical models in economics. *Journal of Economic Methodology, 7*(1), 1–31. https://doi.org/10.1080/135017800362220.

Sugden, R. (2009). Credible worlds, capacities and mechanisms. *Erkenntnis, 70*(1), 3–27. https://doi.org/10.1007/s10670-008-9134-x.

Thicke, M. (2019). Evaluating formal models of science. *Journal for General Philosophy of Science*. https://doi.org/10.1007/s10838-018-9440-1. **(Online version ahead of print)**.

Thoma, J. (2015). The epistemic division of labor revisited. *Philosophy of Science, 82*(3), 454–472. https://doi.org/10.1086/681768.

Weisberg, M. (2006). Robustness analysis. *Philosophy of Science, 73*(5), 730–742. https://doi.org/10.1086/518628.

Weisberg, M. (2013). *Simulation and similarity*. Oxford: Oxford University Press.

Weisberg, M., & Muldoon, R. (2009). Epistemic landscapes and the division of cognitive labor. *Philosophy of Science, 76*(2), 225–252. https://doi.org/10.1086/644786.

Wimsatt, W. C. (1987). False models as means to truer theories. In M. H. Nitecki & A. Hoffman (Eds.), *Neutral models in biology* (pp. 23–55). Oxford: Oxford University Press.

Woodward, J. (2006). Some varieties of robustness. *Journal of Economic Methodology, 13*(2), 219–240. https://doi.org/10.1080/13501780600733376.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In *Proceedings of the sixth international congress on genetic*s (pp. 355–366).

Ylikoski, P., & Aydinonat, N. E. (2014). Understanding with theoretical models. *Journal of Economic Methodology, 21*(1), 19–36. https://doi.org/10.1080/1350178X.2014.886470.

Zollman, K. J. (2007). The communication structure of epistemic communities. *Philosophy of Science, 74*(5), 574–587.