Synthese
https://doi.org/10.1007/s11229-021-03305-2

ORIGINAL RESEARCH

Opacity thought through: on the intransparency of computer simulations

Claus Beisbart[1]

Received: 6 July 2020 / Accepted: 5 July 2021

Abstract
Computer simulations are often claimed to be opaque and thus to lack transparency. But what exactly is the opacity of simulations? This paper aims to answer that question by proposing an explication of opacity. Such an explication is needed, I argue, because the pioneering definition of opacity by P. Humphreys and a recent elaboration by Durán and Formanek are too narrow. While it is true that simulations are opaque in that they include too many computations and thus cannot be checked by hand, this doesn't exhaust what we might want to call the opacity of simulations. I thus make a fresh start with the natural idea that the opacity of a method is its disposition to resist knowledge and understanding. I draw on recent work on understanding and elaborate the idea by a systematic investigation into what type of knowledge and what type of understanding are required if opacity is to be avoided and why the required sort of understanding, in particular, is difficult to achieve. My proposal is that a method is opaque to the degree that it's difficult for humans to know and to understand why its outcomes arise. This proposal allows for a comparison between different methods regarding opacity. It further refers to a kind of epistemic access that is important in scientific work with simulations.

Keywords Epistemology of computer simulation · Understanding why · Verification of simulations · Modeling · Unsurveyability

1 Introduction

Computer simulations are often claimed to be opaque. The rough idea is that computer simulations are not transparent and that their workings are difficult to access. The opacity of simulations is often noted with a sense of regret (e.g., Durán & Formanek, 2018, pp. 645–646) and taken to be relevant for the appraisal of the method

✉ Claus Beisbart
Claus.Beisbart@philo.unibe.ch

[1] Institute of Philosophy, University of Bern, Länggassstrasse 49a, 3012 Bern, Switzerland

of computer simulation. It is thus an important topic in the epistemology of computer simulation (Humphreys, 2009).

Although the opacity of simulations has been noted before (e.g., Di Paolo et al., 2000; Turkle, 1997, 2004), the preoccupation with opacity in the philosophy of science goes back to Paul Humphreys. Drawing on earlier work (Humphreys 1994), Humphreys (2004, pp. 147–151) argues that the relationship between the inputs and the outputs of computer simulations is opaque roughly because the computational steps cannot be known. Humphreys (2009, pp. 618–619) elaborates his claims about opacity to argue for the novelty of computer simulations.

Humphreys's claims about the opacity of simulations have led to a lively philosophical debate. In this debate, the opacity of simulations has rarely been disputed; rather, philosophers have tried to show that opacity doesn't compromise the ability of computer simulations to achieve their tasks. In this vein, Barberousse and Vorms (2014) and Durán and Formanek (2018) have argued that, despite being opaque, computer simulations can produce knowledge. As far as understanding is concerned, Lenhard (2006) has suggested that the opacity of simulation models is compatible with their delivering a pragmatic variety of understanding (see also Lenhard 2009, 2019, ch. 4). Kuorikoski (2011) has discussed how we can improve our understanding of computer simulations despite their opacity. Jebeile (2018) has observed that visualizations can help researchers to use opaque simulations for explanatory purposes. On a more critical level, Saam (2017) has argued that, in the social sciences, opacity is a persistent concern only about one specific kind of simulation. Newman (2016) has suggested that, *pace* Humphreys (2009), the opacity of simulations is not essential and may be avoided by appropriate software construction. Kaminski et al. (2018) have argued that opacity is an exclusive characteristic of simulations only if it is understood in a mathematical sense.

Given this lively debate, it is no surprise that authors have offered additional clarifications of, and reflections on, the notion of opacity. Lenhard (2011) uses the opacity of computer simulations to distinguish them from traditional thought experiments. Imbert (2017, p. 746) lists several ways in which simulations can be opaque (see Imbert 2017, pp. 746–758 for further discussion; cf. Kaminski et al. 2018 for a similar approach). In recent work, Durán and Formanek (2018), Boge and Grünke (forthcoming) and San Pedro (forthcoming) have suggested further clarifications on the opacity of simulations.

Although the discussion so far has led to valuable insights, we still lack a systematic reflection on opacity. As I will argue in detail below, central ideas about opacity have themselves remained opaque (see Kaminski et al., 2018, p. 256 for a similar diagnosis). The term is in fact used differently by different people.

To be clear, opaque objects cannot be seen through. But we can at least think through the opacity of computer simulations. This paper aims to do this. I start (in Sect. 2) with a thorough analysis of Humphreys's influential account of opacity. Since problems for this account will emerge, I consider (in Sect. 3) a recent elaboration in terms of unsurveyability by Durán and Formanek (2018). Section 4 unfolds an alternative approach for explicating opacity. I start from the broader meaning that "opacity" has in ordinary language, and I propose that we think of opacity as a disposition to resist epistemic access, where epistemic access includes both knowledge

and understanding. I elaborate this idea and specify what kinds of knowledge and understanding are relevant for opacity. Here, I can draw on the recent literature on understanding (see Baumberger et al., 2017 for an overview), in particular on Hills (2016). Section 5 offers my conclusions.[1]

Throughout this paper, I restrict myself to computer simulations (*simulations*, for short). This is a method in which a computer is used to trace the time evolution of a system (be it real or merely imagined) by yielding a possibly partial and approximate solution to the dynamical model equations. I illustrate my argument using climate simulations done using the Hadley Centre Coupled Model 3 (HadCM3) by the British Met Office.[2] By the *output* of a simulation, I will mean the 'data' produced by that simulation (e.g., the number 0.44, maybe together with a unit). These 'data' are often visualized using images or animations. When the 'data' are interpreted as descriptions of states within the model implemented, I'll talk of *outcomes*; for instance, one run of a simulation program may yield the outcome that, in the model implemented, the global-average sea-level rise in a certain time span is 0.44 m. I assume that the outcomes include only information that scientists want to learn about the model from running the simulation. I also allow that an outcome is obtained from several model runs and includes estimates of the uncertainties. The *result* of a simulation, finally, summarizes what scientists conclude from the simulation (e.g., that in the real world, a certain emissions scenario leads to a global-average sea-level rise of about 0.44 m in a given period of time).

2 Paul Humphreys on opacity

It's useful to begin with a closer look at Humphreys's pioneering work on opacity.

2.1 Humphreys on the notion of opacity

Humphreys (2004) introduces opacity as follows (pp. 147–148):

> In many computer simulations, the dynamic relationship between the initial and final states of the core simulation [i.e., the process during which the simulation program is run, p. 109] is epistemically opaque because most steps in the process are not open to direct inspection and verification.

Here a computer simulation is conceptualized as a process; its opacity is supposed to derive from the fact that most steps in the process cannot be known.

Humphreys (2009) defines opacity for processes in general as follows:

[1] Opacity and transparency are also discussed in the philosophy of photography (Walton 1984) and in the philosophy of language (Quine, 1953, p. 142). In social epistemology, Wagenknecht (2014) draws a distinction between opaque vs. translucent epistemic dependence among collaborators.

[2] See https://www.metoffice.gov.uk/research/approach/modelling-systems/unified-model/climate-models/hadcm3 for a description (last checked 13.4.2021). I rely on the details published in Gordon et al. (2000) and Pope et al. (2000).

> [(Op-H)] Here a process is epistemically opaque relative to a cognitive agent X at time t just in case X does not know at t all of the epistemically relevant elements of the process (p. 618).

He adds that a process is *essentially* epistemically opaque if all relevant elements *cannot* be known to X due to X's nature (Humphreys, 2009, p. 618).

In more recent work, Alvarado and Humphreys (2017) use the notion of opacity with reference to representations. But since this work does not mention computer simulations, I assume that the 2009 definition reflects Humphreys's mature position on the opacity of simulations (more comments on the 2017 paper follow in Sect. 3.3 below).

2.2 A critical discussion of Humphreys's definition

What should we think of Humphreys's (2009) proposal? Let me first make two observations, and then I will raise an objection.

The first thing to note is that Humphreys's definition does not unfold the ordinary language meaning of "opaque" or "opacity". In ordinary language,[3] "opacity" means the difficulty (i) to look through something or (ii) to understand something (where the second meaning has likely been obtained from the first using a metaphor). When we call *processes* opaque, we do not mean to say that they cannot be looked through in a literal sense; rather, we take them be difficult to look through in a metaphorical sense—because their details are difficult to know. On top, clearly, processes can be difficult to understand. In brief, then, when a process is opaque, there are difficulties with the *epistemic access* to the process. What Humphreys does in his definition of opacity is to focus on a particular aspect of epistemic access: the process is not *known* because not all relevant elements are known. The definition does not refer to *understanding*, but immediately after the introduction of opacity, Humphreys points out that opacity may lead to a loss of understanding (Humphreys, 2004, pp. 148 f.).

Second, Humphreys's preferred notion of opacity is agent-relative. This is natural given that the definition has opacity depend on the agent's knowledge. However, in much talk about opacity (e.g. Humphreys, 2009, p. 621), the notion is not relativized to agents. So there must be a way of interpreting unrelativized occurrences of "opacity". An unrelativized understanding of opacity is indeed useful because philosophers discussing the opacity of simulations are not interested in the relation of individual people to simulations, but rather in features that simulations have generally in relation to people. Humphreys himself calls opacity a feature of computational science (2009, p. 618). A natural way of introducing an unrelativized notion of opacity is to say that opacity *tout court* is opacity to a human being or to a scientist with average-level abilities. Humphreys's notion of essential opacity aligns with this if indeed the agent's nature mentioned in the definition of essential opacity includes average-level abilities of humans. The lack of clarity about what exactly constitutes

[3] See https://www.oxfordlearnersdictionaries.com/definition/english/opacity.

"average-level abilities" does not matter in practice because simulations will turn out to be opaque regardless of how exactly the average level is defined.[4]

My objection is that Humphreys's definition lacks clarity. For one thing, it's not clear what Humphreys means in saying that a person *knows* the epistemically relevant elements. Does she have to know that there are such elements? Or what functions these elements have in the computation? Or what their inputs and outputs are? And whether their results are correct? These questions mention different kinds of knowledge, and the content of the notion of opacity will vary depending on the kind of knowledge that the agent is not supposed to have in instances of opacity.

For another thing, it's not clear what "epistemically relevant elements" are (see Kaminski et al., 2018, p. 265, for a similar criticism, which is soon withdrawn, however, for reasons that I do not really understand). As far as the opacity of processes is concerned, Humphreys (2009, p. 618, fn. 5) clarifies that the answer depends on the type of process. But there is nothing epistemic about types of processes as such. Processes may become the object of different epistemic projects, depending on what the precise aims of the investigation are, and these aims determine what is relevant. It may be objected that the elements have to be epistemically relevant to the agent to whom Humphreys's original definition is relativized. But this won't do because it would follow that many processes are not opaque to an agent simply because she doesn't care about their elements – because she doesn't have an epistemic project that requires knowing the elements.

The problem doesn't disappear when we narrow the focus to the opacity of *computer simulations*. Here, Humphreys thinks of epistemically relevant elements as computations. For instance, Humphreys rephrases the claim that simulations are opaque by saying that "no human can examine and justify every element of the computational processes that produce the output of a computer simulation" (Humphreys, 2009, p. 618).[5] But what exactly are the epistemically relevant elements on this level? Qua computational process, a computer simulation can be split up into elements in several ways (e.g., Barberousse & Vorms, 2014, pp. 3612–3613). According to a very coarse description of the computational process, there is just one computation that provides the approximate solution to some equations. The same process can also be split up into multiple computations, each of which evaluates characteristics such as position or velocity at a time. Under an even more fine-grained description, every call of a function that is pre-defined on the level of the programming language is a computation. Even finer descriptions of the computational steps are possible.

Humphreys does not say what exactly the epistemically relevant computations are. But it's clear that there must be many of them, since the problem is supposed to be that their entirety cannot be known. As just mentioned, Humphreys (2009, p. 618) describes the opacity of simulations by saying that "no human can examine and justify *every* element of the computational processes" (my emphasis). In the

[4] On the basis of this assumption, Boge and Grünke (forthcoming) quantify over agents in their definition of fundamental opacity.

[5] See Imbert (2017, pp. 746, 755) for a similar interpretation of Humphreys's.

terms proposed by Imbert (2017, p. 746), the problem is *global* opacity, while single computations may be transparent (see also his p. 753). Humphreys certainly has an important point because, in some way (given a certain individuation of computations), computer simulations run so many computations that a human agent cannot know them.

But this impossibility leads to opacity in Humphreys's preferred sense only if all computations are epistemically relevant. There is a danger that Humphreys runs into a theoretical dilemma at this point: if, on the one hand, the elements or units are assumed to be relatively comprehensive (e.g., the evaluation of the approximate solution to a set of equations for a specific time), then they seem intuitively relevant, but it's not so difficult to know something about them—for instance, a scientist may know their outcomes, because the elements are few and their outcomes manifest themselves in the output that is examined in detail. Accordingly, the simulation is not opaque to the scientist, contrary to what Humphreys fears. On the other hand, if the units are small (e.g., single additions of numbers), it may be questioned whether each calculation of this sort is epistemically relevant. If this is not the case, then again, opacity does not hold true of simulations. Humphreys himself expresses doubts as to whether finely individuated elements are relevant. As Humphreys (2004, p. 148) explains, abstracting from some details can increase understanding. Humphreys (2009, p. 618) draws our attention to two analogous cases, viz. mathematical proof and scientific instrumentation. If there are philosophically respectable reasons to think that scientists can use instruments without knowing much about their functioning, why not think that a lot of details about computer simulations are likewise irrelevant?

As it stands, then, Humphreys's definition lacks clarity: it's not clear what Humphreys means in saying that an agent must know the epistemically relevant elements of the computer simulations. The reply that it all depends on the agent and her epistemic projects won't do because it renders opacity a matter of arbitrary interests. Fortunately, Durán and Formanek (2018) have recently made a proposal to clarify Humphreys's definition.

3 An elaboration of Humphreys's definition: opacity as lack of surveyability

3.1 The elaboration explained

Durán and Formanek (2018, p. 615) elaborate Humphreys's (2009) definition as follows:

> [A] process is epistemically opaque relative to a cognitive agent X at time t just in case X at t doesn't have access to and can't survey all of the steps of the justification.

The process of a computer simulation is now conceptualized as justification and regarded as an argument: descriptions of a series of states are inferred from the specification of initial conditions and the model assumptions (cf. Beisbart, 2012).

This process yields at least conditional inferential justification: the state descriptions are justified conditional on the initial conditions and the model assumptions. The justification is subject to a normative standard, viz. validity: the state descriptions have to follow from the initial conditions and the model assumptions. This standard is fulfilled if, and only if, the computational processes are correct in the sense that they provide approximate solutions to the equations of the model that the researchers have intended to implement in the simulations. To simplify the terminology, let's just talk of the *correctness* of the simulations when referring to the equivalent normative standards just mentioned.

To be sure, many simulations are subject to a stronger standard since they are supposed to provide accurate descriptions of a real-world target system. Accordingly, such simulations are supposed to justify the descriptions in an unconditional way. However, there are other simulations that do not have a real-world target system because they are only supposed to trace the behavior of a model, for instance, a world in which point particles attract each other with a modified version of the gravitational force. So a general account of opacity cannot assume that this stronger, unconditional justification is needed, and the account has to be restricted to the weak, conditional form of justification mentioned above.[6]

In an opaque simulation, the authors take the justification/computation to be *unsurveyable*. The notion of unsurveyability is borrowed from Tymoczko, who calls a mathematical proof surveyable if, and only if, it "can be definitively checked by members of the mathematical community" by hand (Tymoczko, 1979, pp. 59–60). According to Tymoczko, Appel and Haken's proof of the four-color theorem is not surveyable since a computer had to be used to show that more than 1′000 configurations each have a property called reducibility. A human being cannot show this in a reasonable amount of time.[7]

By extending Tymoczko's idea of unsurveyability from proofs to all arguments, we can say that arguments are unsurveyable if their validity cannot be checked by hand. Accordingly, we can summarize the elaboration by Durán and Formanek (2018) as follows:

> (Op-DF) Processes of justification (in particular computer simulations) are opaque if, and only if, their correctness cannot be checked by hand.

Most computer simulations are opaque in this sense. A human agent cannot immediately see or comprehend that the outcome has been correctly derived from

[6] Durán and Formanek (2018) overlook this point. As a consequence, they take validation to offer a solution to the problem of what they consider to be epistemic opacity. The task of validation is to show that the results of computer simulations reflect their target appropriately. For showing this, it is often important to make a case that the simulation solves the equations correctly (Beisbart, 2019), which is to make a case that the conditional justification works. But whether the computer simulation and the underlying model also represent the target appropriately, does not much impact on opacity, as defined by Durán and Formanek.

[7] Teller (1980) objects that Tymoczko mischaracterizes the proof of the theorem. For Teller, the authors of the proof did survey all configurations, albeit with the help of a computer. As we will see, a similar standpoint is possible regarding computer simulations.

the input and the equations constitutive of the model. Nor is it possible to split the simulation into steps each of which can be checked by a human by hand. In this way, the proposal escapes the dilemma described in Sect. 2.2 above. The reason is that *every* segmentation of the computational process leads to a situation in which the whole computation cannot be checked for correctness. If we call elements of the process "epistemically relevant" if, and only if, their correctness can be checked by hand, we can say that the opacity of simulations is due to the large number of epistemically relevant steps—as Humphreys claims.

The elaboration also clarifies what "knowing the steps" means: checking the steps for correctness. So the unclarities noted above have been removed. There is also textual evidence that the elaboration articulates what Humphreys had in mind. For instance, when Humphreys (2009, p. 618) glosses his claim that simulations are opaque by saying that "no human can examine and justify every element of the computational processes," he refers to justification.

But how convincing is the elaboration from a systematic point of view?

3.2 Problems with the elaboration

A first problem is that opacity à la Durán and Formanek merely articulates the challenge of so-called verification, a challenge well-known in the literature. Let me explain.

The authors reduce opacity to the impossibility of checking the argument implicit in the simulation. This argument is supposed to derive the outcomes of the simulation from the model assumptions, the assumed initial conditions, etc. To check that the simulation has in fact traced the consequences of these assumptions is the task of so-called verification. As it is often put, verification is about "solving the equations right" (e.g., Roache, 1997, p. 124, who refers to other authors).[8] That a simulation is opaque à la Durán and Formanek then means that verification cannot be done by hand. But this is very well-known. Precisely for this reason, verification is typically done using tests that check whether the program reproduces computations with known outcomes (i.e., known solutions to the model equations). There is an extensive literature on how exactly this should be checked and what the difficulties are (see, e.g., de Millo et al., 1979; Oreskes et al., 1994; Roache, 2019).

A second problem arises due to the general strategy that Durán and Formanek adopt, following Humphreys. The strategy is, broadly speaking, not to content oneself with the idea that opacity is the disposition to resist epistemic access, as the dictionary definition would suggest. Rather, the strategy is to "dig deeper" and to point to a specific way in which epistemic access to computer simulations is difficult. Very likely, the aspiration is to explain why and how computer simulations resist epistemic access.

To seek such an explanation is worthwhile. But there is no reason to absorb such an explanation into the very notion of opacity. An "etiological" definition of opacity

[8] See also Schlesinger et al. (1979) and Oberkampf (2019, pp. 70, 75–79) for the notion of verification.

would make sense if there was one single way in which epistemic access is difficult for humans. But this condition is not met. As will become clear below, computer simulations are difficult to access for several reasons. A definition that mentions one salient aspect is problematic because it is likely to miss others and to be too narrow if compared to the ordinary language meaning of "opacity."

To be more specific: already Humphreys's definition concentrates on lack of knowledge of epistemically relevant elements as a salient cause. This is too restrictive because lack of knowledge is only one way in which epistemic access to computer simulations may be limited or prevented. It may well be the case, for instance, that humans struggle with computer simulations at least in part because they do not fully *understand* them, even though they know a lot about them. Humphreys further focuses on knowledge of the relevant *computational* steps. Computer simulations cannot only be described on the computational level, but also using other layers, e.g., the physical one (Barberousse et al., 2009), raising the question of why Humphreys considers only knowledge on the computational level. The Durán and Formanek elaboration (Op-DF) is, maybe, not restricted to one specific layer. But it's clearly focused on a particular type of epistemic access to computer simulations.

To see why the difficult epistemic access to computer simulations is not exhausted by the features that Humphreys and Durán and Formanek point to, consider the following thought experiment. Suppose that super-scientist Susan knows all the relevant details about the huge number of computations that occur during a computer simulation and that she has checked the simulation qua justification. There are still two things preventing the achievement of full transparency. First, even if Susan knows all the required information about the computations, she doesn't necessarily grasp the connections between the computations. It is one thing to know that, say, in step t_n, variable v has taken a value larger than 1, and another thing to grasp how this is connected to earlier steps. The reason why the value of v is larger than 1 may be that, in step t_{n-1}, v had a value larger than 0.8 and that two other variables each had values larger than 0.1. Susan knows that, in step t_{n-1}, v had a value larger than 0.8 and so on, but this knowledge does not imply that this is the reason for why the value of v is larger than 1 one step later. Nor does the fact that Susan has checked every step imply this. Second, Susan may still not be able to engage in counterfactual reasoning about the simulations or computations. For instance, what would happen if a particular epistemically relevant computation got it wrong due to a specific mistake? Neither knowing the actual results of the epistemically relevant computations nor having checked the calculations (nor the two combined) enables Susan to run the inference about the counterfactual situation. The counterfactual inference requires the ability to do something else, for instance, calculate the results in order to find out what is implied in the counterfactual scenario, or anticipate the results due to an intimate knowledge of the model.

What is lacking here is ultimately understanding. For it is a commonplace in the literature about understanding that the latter requires grasp of connections (e.g., Kvanvig, 2003, p. 192; Grimm, 2011, p. 88) and the ability to run inferences about actual and counterfactual scenarios with slight variations (e.g., Grimm, 2006, pp.

532–533; Hills, 2016, p. 3).[9] What the thought experiment shows, then, is that the notions of opacity discussed so far are indeed narrower than what we would call the opacity of simulations in ordinary language.

It may be objected that appeal to a broader notion of opacity from ordinary language is misguided because ordinary language is irrelevant to the epistemology of computer simulations. The task of the latter is to identify the potentials and possible pitfalls of computer simulations, but not to provide a conceptual analysis of the term "opaque"—or so the objection goes. In response, I grant that we are not in the business of conceptual analysis. But what's the point of calling a special idea about computer simulations "opacity" if there is no relation to opacity, as it is understood in ordinary language? On a charitable reading, the task that Humphreys etc. have set themselves is giving an explication of the notion of opacity for computer simulations (see Carnap, 1950/62, ch. 1). One desideratum for such explications is similarity with the relevant notion from ordinary language.

In fact, if we call a very specific feature of simulations "opacity," then we are likely to slip back to the ordinary language notion of opacity. This has indeed happened in the literature. Imbert (2017) argues that there are various sorts and origins of opacity of simulations (p. 746). One such sort of opacity is the opacity defined by Humphreys, but Imbert points to what he takes to be another variety of opacity, which is supposed to derive from the fact that computer simulations are often run by groups of experts coming from different fields. The thought here is not so much that there are too many epistemically relevant elements (which would have to be the crucial point if Imbert was talking about opacity as defined by Humphreys). Rather, the problem is that expertise in one field does not suffice for understanding the whole simulation. Thus, Imbert cannot be referring to opacity as defined by Humphreys.

It's interesting to note that the recent definition of opacity by Alvarado and Humphreys (2017) moves closer to the ordinary meaning of opacity. According to them, a representation counts as opaque if, and only if, it does not represent "the states of a system in a way that is open to explicit scrutiny, analysis, interpretation, and understanding by humans" or if it is not the case that "transitions between those states are represented by rules that have similar properties" (p. 740).

Here, "explicit scrutiny, analysis, interpretation, and understanding by humans" are the sorts of things that I have called epistemic access. But for our purposes, this definition will not do. First, it's not clear how the definition, which is concerned with opaque representations, applies to computer simulations. While computer simulations clearly involve representations, it is not clear whether they *are* representations. Further, simulations involve various layers of representations (e.g. Küppers & Lenhard, 2005; Winsberg, 1999). The question then is at which layer we should apply the new definition. Second, there are clear cases of simulations in which a central layer of representation, viz. the conceptual model, is not opaque in the sense

[9] Sometimes, the ability to run inferences about counterfactual scenarios is assumed to be part of grasping connections (see, e.g., Grimm, 2006, pp. 532–533). Whether or not this is so, it does not matter for our argument as long there is an ingredient of understanding that does not reduce to knowledge about the calculations.

defined here. For instance, the state descriptions and the transition rules in Conway's game of life (Gardner, 1970; Wolfram, 2002) are very simple and thus accessible. But, intuitively speaking, related simulations are nevertheless opaque because it is very difficult for humans to obtain an understanding of how their outcomes arise.

4 A new proposal: opacity as disposition to resist epistemic access

Can we do better than Humphreys or Durán and Formanek? In this section, I make a fresh start and begin with the idea that opacity is just what the dictionary suggests: something is opaque if, and only if, it is difficult to know (i.e., difficult to see through in a metaphorical sense) or difficult to understand. As before, I summarize this by saying that opacity is a disposition to resist epistemic access by humans. I apply this idea to computer simulations and inquire more systematically into the ways they are difficult to access.

This difficulty in being known and understood is naturally conceptualized as a disposition, so I'll explicate opacity as a disposition. To be sure, this choice is not without alternatives. Humphreys defines opacity simply as lack of knowledge and understanding. But such a definition of opacity doesn't get us to the core of opacity. An agent might not know or understand something, but unless there is a barrier to her knowing or understanding that thing, opacity cannot be the culprit. As an alternative, opacity may be defined modally as the impossibility of having suitable epistemic access (this is the idea behind Humphreys's definition of essential opacity). But this yields the unwelcome consequence that there is no way to overcome opacity.

Our idea to start with the ordinary meaning of opacity needs more work because many things about a simulation may be known or understood. By contrast, simulations are not opaque just because they contain details here and there that people may want to know and understand, and that turn out to be difficult to know and understand. Lack of knowledge or understanding of some details is of interest only if the details are—well, relevant. What then are the relevant things that we need to know and understand to avoid opacity?[10]

This question has arisen before in the discussion of Humphreys's definition. For an answer, it suffices to specify certain *types* of things that are interesting, and which turn out to be hard to know or to understand.

4.1 What knowledge and understanding are relevant?

Relevance is a relation, so to what do knowledge and understanding have to be relevant? A natural answer is that knowledge and understanding must be relevant to the

[10] Note that I can concentrate on the understanding *of* a computer simulation and bracket the understanding via computer simulation. For the question of how computer simulations may be used to understand a target system, see Fernández (2003), Lenhard (2006), Kuorikoski (2011), Grüne-Yanoff (2009), Ylikoski (2014) and Parker (2014).

primary goal for which a given computer simulation is run—where the *primary* goal is the most subordinate goal for which the simulation as a whole is done. Superordinate goals for the sake of which the primary goal is undertaken (to improve one's model, for instance) vary and need not be considered in a general investigation into the opacity of simulations.

For each single run of a simulation program (the unit on which I will focus), the primary goal is to obtain information about a dynamical evolution that unfolds in a model given specific initial conditions (cf. El Skaf & Imbert, 2013). This information is contained in state descriptions that can be obtained from the output. Often, the primary goal is more specific and focused on certain aspects of the state descriptions, say, a certain pattern of precipitation in some region or its projected average temperature for the next ten years with a specified accuracy rating. The information that a simulation yields in relation to its primary goal is what I call the *outcome* of the simulation. I assume that the outcome can be formulated in terms of a proposition p about the sequence of states in the model investigated by the simulation. For an example, consider what Gregory and Lowe (2000, p. 3069) report in an application of the HadCM3 model: "global-average sea-level rise from 1990 to 2100 is predicted to be […] 0.44 m in HadCM3."

For reasons of simplicity, I will often assume that the outcome is only about a final state at some final time t_f. The generalization to a series of times is straightforward.

A natural question to ask then is this: why did a specific outcome arise from the simulation? For short, why p? Every answer to this question will give us knowledge of why p. Furthermore, there is the task of understanding why p. If Hills (2016) and others are right, then understanding why p goes beyond merely knowing why p. Instead, understanding why p requires an agent to grasp the relation between p and its explanation: call it q (Hills 2016, p. 3).

It is thus natural to propose that the knowledge and understanding relevant to opacity are:

- knowledge of why p
- understanding why p

for the outcomes of simulations p.[11] But what exactly does this knowledge and understanding amount to and how difficult are they to obtain?

a. Knowledge of why p

To know why a specific outcome (p) has arisen is to know an explanation of why p has arisen. Every such explanation will involve the computer simulation as process. This process has various layers that can be described using different vocabularies (Barberousse et al., 2009). For the purposes of this paper, we can distinguish between three layers: The *physical* layer is described by referring to the hardware in which physical entities such as wires or electrons interact with

[11] Boge and Grünke (forthcoming) go in a similar direction when they require insight into the way the output is obtained from the input.

each other. In the *computational* layer, the process is interpreted as doing computations (e.g., as multiplying two numbers as a contribution to calculating partial solutions to certain equations). Regarding the third, *representational* layer, the process is described using a sequence of states in the model (e.g., a series of states of the climate of a model of the Earth).

There are thus basically three kinds of explanations of p, call them q_j (explanations situated on different levels or combining the layers in different ways need not be considered because they are even more complicated). The *first* explanation q_1 operates mostly on the *physical* level; the final physical state of the computer simulation is explained using physical laws for the hardware (e.g., Kirchhoff's circuit laws) and an earlier state of the hardware—typically the initial state of the simulations. Finally, the description of the final physical state of the hardware is translated into a state description of the model following the conventions underlying the simulation. This step is needed because the outcome p is cast in terms of the model evaluated by the simulation.

The *second* explanation q_2 is mostly *computational*; the final computational state of the simulation is explained in terms of the computations done during the simulation following the algorithm, and an earlier computational state, naturally at the beginning of the simulation. This explanation assumes that the computations are properly done by the hardware and that there are, for example, no hardware failures that lead to deviations from the application of the algorithm. The final computational state then is translated into a state description at the representational layer, in a similar way as in the first explanation.

Finally, the *third* explanation q_3 is mostly on the *representational level*. The final state in the model is explained using the dynamics holding in the model and the initial conditions set within the model. In the coupled climate model HadCM3, for instance, there are two components, viz. the atmosphere and the ocean. The states of the components are coupled once per simulated day to describe the interactions (Gordon et al., 2000). The explanation of the outcomes at this level has to assume that the computer simulation traces the model as intended and that there are, for example, no approximation errors that lead to a deviation from the model implications.

Thus, each explanation is an inference that leads from a description of the initial state of the simulation and a catalogue of suitable dynamical laws on a particular level to a statement about the final state of the simulation. The physical and computational explanations need, in addition, a translation of the description of the final state into the terms of the model because the phenomenon to be explained is a model state (the outcome). Further, the computational and representational explanations must include a premise to the effect that processes on the lower levels work as intended.[12]

To know why p, the agent needs to know at least one explanation of p; and to know this, she needs to know at least the premises of the explanation and, maybe,

[12] My account doesn't assume the deductive-nomological (DN) account of explanation, but only the idea that explanations are arguments or inferences.

that the premises explain or imply the outcome. She need not be able to run the related inference on her own—she may know the explanation from testimony, for instance.

The first two explanations, the physical one (q_1) and the computational one (q_2), are too difficult for a human being to know. Knowing the physical explanation would require detailed knowledge of the computer and its dynamics because the outcomes depend on tiny details of the physical structure. Knowing the computational explanation would demand detailed knowledge of complicated computations that have been arranged according to the algorithm.[13] Only the third, representational explanation q_3 can realistically be known to humans. To know it, an agent needs to know the premises of the explanation, that is: the initial state of the model ($q_3{}^a$), the dynamics of the model ($q_3{}^b$), and that the simulation works as intended ($q_3{}^c$)—in other words, the agent needs to know that the simulation solves the model equations to a sufficient approximation.[14] Two of these, $q_3{}^a$ and $q_3{}^b$, are easy to know. Knowledge of $q_3{}^c$ is usually established during verification. Verification is difficult, but often a sufficiently strong case can be made that the simulation solves the equations correctly (see Rider, 2019). Finally, maybe, the agent needs to know that $q_3{}^a$ through $q_3{}^c$ imply p. This knowledge is easy to have too; what is required beyond knowledge of $q_3{}^a$ through $q_3{}^c$ is only the observation that the computer simulation has p as its outcome.

All in all, using explanation q_3, knowledge of why p is not too difficult to acquire.

b. Understanding why p

Let us turn now to understanding why p. A useful proposal on what this amounts to is given by Hills (2016). Following the idea that understanding why p requires grasping the relationship between p and an explanation of it, q, Hills (2016, p. 3) specifies the following abilities as conditions on understanding why p:

i. follow some explanation of why p given by someone else.
ii. explain why p in your own words.
iii. draw the conclusion that p (or that probably p) from the information that q.
iv. draw the conclusion that p* (or that probably p*) from the information that q* (where p* and q* are similar to but not identical to p and q).
v. given the information that p, give the right explanation, q.
vi. given the information that p*, give the right explanation, q*.

[13] Of course, scientists may have a rough knowledge of the explanation by knowing that the simulation has approximately solved this and this equation. But this is not to know a detailed explanation of the outcome at the computational level.

[14] Things get more complicated for simulations that don't work as intended, but we can bracket such simulations because if simulations that work are difficult to understand, so will be simulations that don't work.

Given knowledge of why p, the abilities i, ii, iii, and v are not very interesting. They mean only that an agent can work with her knowledge of why p in a suitable way—that she can mention this knowledge if asked, etc.

For our purposes, abilities iv and vi are more interesting. They require that the agent can run inferences about actual or counterfactual scenarios similar to the ones investigated in the computer simulation. More specifically, iv. requires the agent to infer variations p* of p from variations q* of q, while vi. mainly requires the agent to run inferences from variations p* of p to variations q* of q. Such abilities are also deemed central for understanding in the context of simulations by Kuorikoski (2011). The abilities iv. and vi. also cover the ability to infer which parts of the model are relevant for the outcome, an ability that Imbert (2017, p. 754) stresses in relation to opacity (see also Jebeile, 2018, p. 214; in her terms, there is a gap between the model and the outcome of a simulation).

Abilities iv and vi are both extremely demanding. For an illustration, consider again the outcome reported by Gregory and Lowe (2000, p. 3069): "global-average sea-level rise from 1990 to 2100 is predicted to be […] 0.44 m in HadCM3." In this case, variations q* of the explanans q can arise by changing a. the initial conditions; b. the model assumptions, in particular the model dynamics (e.g. the cloud scheme used in the description of the dynamics of the atmosphere); or c. the degree to which a computer simulation works as intended. Under each type of modification of q, it is extremely difficult to infer how p would change. As an example, focus on b. the model assumptions (e.g., the cloud scheme): the question is what would happen if the assumptions implicit in the cloud scheme were changed. Suitable inferences may, in principle, be run on any of the three levels. However, inferences on the level of the model are too difficult because the model equations cannot be solved analytically. A numerical approximation is also very difficult because either the single steps are too difficult or there are too many of them. Inferences on the level of the computations are difficult for precisely the same reason. Inferences on the level of the computer hardware, finally, are difficult because the hardware processes are too complicated.

Similar results apply to inferences from p* to q*. To have ability vi, scientists would have to be able to infer how (for instance) the initial conditions or the cloud scheme might be changed to obtain a different sea-level rise (of, say, 0.32 m instead of 0.44 m).

The upshot, then, is that understanding why the computer simulation has yielded a specific outcome (p) is extremely difficult because this would require the ability to run inferences between variations on the setup of the simulation and its outcome. According to my proposal, this understanding is relevant to an assessment about opacity, so computer simulations are opaque.

In this explanation of why simulations are opaque, the vast number of computations (the factor stressed by Humphreys) plays a role because it makes grasp of the explanatory connections on the computational level impossible. But this grasp is not strictly necessary for understanding why p: if an agent were able to run the required inferences about modifications of p and q on the physical or on the representational level reliably, then the simulation would not be opaque to her anymore, according to my definition. The fact that a grasp of the physical level and the model behavior are

extremely difficult for humans adds to the full explanation of why, according to my proposal, simulations are opaque.

It may be objected that scientists can run the required inferences about counterfactual scenarios by running variations of the computer simulation under scrutiny. A slight variation of the initial conditions can be studied by running the simulation program with a modified input. This is in fact what climate scientists do when they run initial condition ensembles (see Parker, 2010 for a discussion of various kinds of ensembles). Alternative variations that refer to the model assumptions or to the working of the computer simulation may need some changes in the simulation program, but using the modified program, the simulation scientists can infer what would happen under the variations (see Kuorikoski, 2011, Sect. 4.2 for this perspective). Again, this is something that climate scientists do. For instance, Gregory and Lowe (2000) compare two different versions of the Hadley Centre Coupled Model regarding global sea level. Collins et al. (2007) study how the outcomes of their simulations are affected if certain parameter values are changed, for instance, the diffusivity of tracers along certain surfaces in the ocean. They explicitly say that they do this in order "to understand the leading-order impact of the perturbations on future climate change" (Collins et al. 2007, p. 2316). In their final passage, they state that "[t]he next stage of the work will be to understand why the simulations show only small changes in ocean heat uptake efficiency and global mean temperature change" (Collins et al., 2007, p. 2320). This is precisely the kind of understanding that I argue is important.

True, running a second (modified) computer simulation can in principle help a scientist to run an inference about a counterfactual scenario. But this option doesn't mean that the opacity (in the intended sense) of computer simulations is reduced. Already one run of a computer simulation that is needed to make a counterfactual inference will typically take a considerable amount of CPU time—coupled climate models that trace the climate for 400 years with a time step of 30 min (Gordon et al., 2000) take a lot of CPU time.[15] This CPU time may not be available at all, in which case the scientist cannot run the required inference in practice, so she lacks an ability constitutive of understanding. If, alternatively, the CPU time is available to her, running the simulation still takes some time, so the scientist cannot be said to run the inference immediately. Note here that our notion of opacity doesn't require that the inference is impossible. It's sufficient for opacity that such an inference is difficult for agent, and this is the case if a lot of CPU time must be invested.

But perhaps a different objection can be leveled against my claim that scientists have a hard time understanding why the outcomes of a simulation have arisen. This objection holds that scientists can easily learn to run the required counterfactual inferences by running the simulation program several times and checking the outputs using visualizations. On this basis, they are trained to run the counterfactual inferences (Lenhard, 2019, p. 100 claims that this method can provide orientation within the model). They may either do this intuitively (in which case we may want

[15] According to Hanappe et al. (2011, Table 1), simulating one month using a low-resolution version of HadCM3 on a CELL's PPE takes about 4 h CPU time.

to speak of tacit knowledge or expert judgement) or because they have learned rules for running the inferences. If it is correct that scientists can easily learn this, it seems to follow that computer simulations are not, on my proposal, very opaque at all.

The method that Lenhard describes is certainly helpful in overcoming opacity. But the fact that there is this method doesn't show that the simulations are not opaque to begin with. One reason is that the method is computationally very expensive. For many simulation runs that each result in a specific outcome, several other runs are needed to trace "close" counterfactual scenarios—this in order to obtain a sufficient basis to run further inferences. But if the method is computationally costly, then it is not easily used. A second reason is that the power of the method is limited. For one thing, most often, the method will only enable a scientist to run inferences about counterfactual scenarios in a qualitative way. For instance, the scientist will be able to infer that precipitation increases if a certain other parameter is increased. But very often, quantitative predictions are of interest, and this qualitative reasoning will not allow for sufficiently precise inferences. For another thing, as described, the method is restricted to the consideration of changes within the model. But what is also needed for a fuller understanding is the ability to run inferences about scenarios in which e.g. the hardware doesn't work as intended or the implementation of the method is changed. Of course, the method could be expanded to cover such counterfactuals, but the consequence would be that things get even more expensive.

The objections we've just considered assume that the agent can run further computer simulations to increase their understanding of why the results have arisen. It may be responded that we must not appeal to further work that computer simulations can do if we discuss the opacity of simulations. Our topic is the relationship between humans and simulations, and this requires that we keep the relata separate. It would seem strange to shift computer simulations to the side of ourselves, as it were, and to allow that humans use them as tools, when we consider this relation—or so goes the response. This response has some plausibility, but it's not clear whether it is fully adequate. In Sect. 3.2, we have allowed that scientists check a simulation program by running it for verification, so why not allow that scientists run the program to improve their understanding? The deeper issue in the background is how we understand ourselves in relation to machines. As suggested by Clark and Chalmers (1998) or Humphreys (2004), the view that humans alone are the epistemic subjects may have become outdated or inappropriate; today, coupled systems that include machines may be the relevant agents. This issue cannot be resolved in this paper; nor can we discuss what exactly opacity means if the agent is allowed to be a coupled system. For this reason, I have to leave open whether the objections can be countered using this response. This is unproblematic because I have offered different reasons to reject the objections.

4.2 A new explication of opacity

I can now condense my suggestions into an explication of opacity. The general idea is that some X is opaque if, and only if, it's difficult, if not impossible, to know and to understand why the outcomes of X arise. This idea can be applied to all methods

(cf. Kaminski et al., 2018, p. 258), so I'll define the opacity of methods. Doing so is an added value because opacity may be a concern about other methods too (e.g., accelerator experiments in particle physics).

I assume that each method produces outcomes that can be cast in terms of propositions. For instance, for experiments, the outcomes specify what is observed or measured about the system on which the experiment is done. Since a method allows for many applications that each have a specific outcome, I first define what it means that a specific application of a method is opaque, and then generalize to the method. When doing so, I need not assume that all applications of an opaque method are opaque. This is a plus because a simulation program could (in some instances) be run on a set of initial conditions for which an analytic solution is available; in such a case, understanding the outcome would not be so difficult, because the solution would facilitate reasoning on counterfactual scenarios.

A question that the explication still has to answer is to whom the outcomes of applying a method are difficult to know and understand. A natural proposal refers to the following situation: Scientists have constructed a setting in which the method can be applied (e.g., they have built up an experiment, programmed a computer simulation, etc.) and applied the method for the first time. In the case of a simulation, this means that they know the variables, the model that is implemented, and details about the implementation. If they nevertheless have difficulties knowing and understanding why the outcomes have arisen, there is opacity. Thus, let us say that we are interested in the difficulty for *average scientists in the default situation*. Since the level of competence of this group of scientists is very high, the difficulties with epistemic access extend to other groups, notably including laypeople. Even scientists who know and understand to some extent why the outcome has arisen have acquired this knowledge and understanding only with difficulty. What we have to exclude though is a scenario in which average scientists in the default setting have significantly improved their understanding of the simulation and the underlying model, perhaps by running the simulation very often. But scientists who have done this have hardly any motivation to fulfill the primary goal of a simulation, viz. to learn the outcome for a specific initial condition.[16]

One further preliminary note before I specify my explication: Understanding, which figures centrally in my explication, comes in degrees (see Baumberger, 2019). This is presupposed by ordinary talk, for instance, when we speak of better or deeper understanding. Note also that the abilities that have been taken as central for understanding (following Hills) may be possessed to various degrees. For instance, an agent may be able to run some, but not all of the required inferences (see Hills, 2016, pp. 4–7 for a detailed discussion of the degrees of understanding why). But if understanding, and thus a central component in the explication is gradable, so should be the explicatum. There are indeed good reasons to think that opacity is gradable, at least in its second ordinary language sense, that is, qua being difficult to understand. I'll thus define a gradable notion of opacity

[16] Knowledge of the variables, the model assumptions, and the implementation form part of a broader objectual understanding of the simulation. But this broad understanding is not our focus here.

(as does San Pedro forthcoming; but he stays closer to Humphreys's original definition). Using this gradable notion, we can naturally introduce an ungradable one by saying that something is opaque full stop if, and only if, it is more opaque than a contextually fixed standard.

Here then is my explication of opacity:

1. The application of a method is opaque to the extent to which it is difficult for average scientists in the default setting to know and to understand why the outcome has arisen.
2. A method is opaque to the extent to which its typical applications are opaque.

A few comments are in order to explain this proposal.

First, the explication mentions both knowledge of why p and understanding why p. It is very plausible to think that the latter is not included in the former. Some authors, notably Kvanvig (2003, pp. 197–200), Pritchard (2010, p. 13), and Hills (2016, pp. 11–18), have likewise argued that understanding why p does not imply knowing why p. In the explication, I remain on the safe side if I include both knowledge and understanding why p.

Second, the degree to which a method resists knowledge may differ from the degree to which it resists understanding; in those cases, the degree to which *knowing and understanding why p* is difficult is meant to identify the degree to which the joint task constituted by knowing and understanding why p is difficult. The idea here is that knowledge and understanding form a package and that we are talking about the difficulty of obtaining the whole package. In most cases, understanding why p is more demanding and more of a concern.

Third, the proposal does not presuppose any specific account of knowledge and understanding (or of explanatory understanding, more precisely, see, e.g., Baumberger 2011, pp. 70–71). This is as it should be, because the ordinary language term "opaque" is naturally explained in terms of knowledge and understanding and because a philosophical clarification should not depend on a particular account of (for instance) understanding. Above, I have used Hills's account of understanding why p to show that computer simulations are opaque. What was crucial for my argument that computer simulations are opaque was the assumption that understanding why p requires the ability to run inferences that lead from slight modifications q* of q to modifications p* of p and back. This sort of requirement is not unique to Hills; it is also required by (e.g.) Grimm (2006, pp. 532–533). Independent of a philosophical account, it is very plausible to say that it's difficult to understand why the outcomes of a simulation are such and such.

Fourth, the typical applications of a method are restricted to the intended applications. The latter vary with respect to (e.g.) initial conditions and parameter values, but for both types of choices there are natural measures; for instance, a Lebesgue measure for a realistic range of parameter values. Typicality then excludes applications with a high symmetry that often have zero Lebesgue measure.

My explication of opacity has several advantages over Durán and Formanek's proposal.

It is, first, broader than theirs. Their opacity contributes to opacity in my sense, but mine runs deeper. This is because unsurveyability makes explanations and understanding related to the computational level very difficult, but, according to my proposal, opacity is not just constituted by unsurveyability; rather, the core is that scientists do not understand why the results arise. As shown with the super-scientist Susan above, a scientist who can check all the computations in a computer simulation may nevertheless be at loss to tell why the results arise. Further, my proposal takes the physical and representational levels to be important as well. Consequently, good understanding of a simulation on either level may help to overcome opacity. That my proposal is broader in this sense is a plus because it covers more of the epistemological concerns that we may have about computer simulations.

Second, my explication is compatible with the idea that there are various aspects to opacity (this is a central claim by Kaminski et al., 2018). Since the possible explanations of why the outcome has arisen can be situated on different levels, the levels can be taken to define several aspects of opacity. Likewise, my explication can make sense of Imbert's (2017) idea that there are various sources of opacity.

As a third advantage, my proposal has broader scope in that it applies to all methods. Durán and Formanek's proposal, by contrast, is restricted to justifications. But some methods do not justify their outcomes. For instance, an experiment does not justify its outcome (what may need justification is that a certain proposition is indeed part of the outcome).

Finally, it's also an advantage that my proposal is closer to ordinary language. At least sometimes when researchers talk about opacity, they use the term in the sense known from ordinary language, for instance, when Di Paolo et al. (2000, p. 497) say, of the opacity of simulations: "[I]t is not immediately obvious what is going on or why."

5 Conclusions

Although computer simulations are often claimed to be opaque, it's not entirely clear what exactly the opacity of simulation amounts to. In this paper, I have argued that Humphreys's pioneering account of opacity is a bit narrow since it concentrates on the huge number of computational steps. I have thus proposed a more comprehensive notion of opacity by starting with the ordinary language meaning of opacity, viz., that something resists epistemic access. Appeal to ordinary language is fully legitimate at this point because scientists and philosophers are likely to slide back to the ordinary meaning of the term.

To make sense of this option, I had to explain which type of knowledge and which type of understanding are relevant to opacity. According to my proposal, a method is opaque to the extent to which it is difficult to know and understand why the outcomes of the method arise. The question of why the outcomes arise is a very natural one, given that simulations have the primary goal of obtaining the outcome as information about a model. Drawing on Hills's account of understanding why p, I have argued that what makes simulations opaque is this: we cannot run inferences about slight modifications to the setting of the simulation and to its outcome.

A worry may be that the concept of understanding is too vague to help to clarify opacity. It may also be feared that the various authors I've drawn on (e.g., Baumberger, Hills and Lenhard) ultimately hold incompatible views about understanding. However, the recent literature on understanding has clarified that there are different varieties of understanding. For instance, it is pretty uncontroversial that objectual and explanatory understanding can be distinguished. What we crucially need for the purposes of this paper is a notion of explanatory understanding that goes beyond mere knowledge of explanations; this is not completely uncontroversial (see Baumberger et al., 2017, end of Sect. 4.2 for an overview), but many authors would agree. The views that Baumberger, Hills, and Lenhard adopt are fully compatible. While Baumberger (2019) is concerned with objectual understanding, Hills focuses on explanatory understanding. Baumberger's point that understanding comes in degrees is admitted by Hills. Lenhard's (2006) primary aim is to propose a notion of pragmatic understanding that is compatible with opacity, but different from more traditional ideas about understanding.

My proposed account has several virtues, or so I have argued. It can be applied to all kinds of scientific methods, including, for instance, machine learning. In this context, it's promising that my account fits extremely well what Burrell (2016, p. 1) says about the opacity of machine learning:

> [The machine learning algorithms] are opaque in the sense that if one is a recipient of the output of the algorithm (the classification decision), rarely does one have any concrete sense of how or why a particular classification has been arrived at from inputs.

Since the proposed notion of opacity is gradable, it can make sense of the claim that machine learning is even more opaque than computer simulation (see Symons & Alvarado, 2016 for work that discusses opacity in relation to big data). To what extent this claim is in fact true is an important topic for future research (see Boge and Grünke forthcoming for pioneering work in this direction).

What is more important than the conceptual question of what exactly we should mean by "opacity" is an analysis of the epistemic problems that computer simulations raise. In this regard, my paper has drawn attention to the fact that it's very difficult to explain and to understand why the outcomes of a specific computer simulation arise. Lack of this kind of epistemic access to simulations hampers the work researchers may hope to carry out with simulations. Understanding why a certain outcome has arisen can help during the verification of simulations, for instance, to locate a bug. Further, if scientists understand how an outcome has arisen in a simulation, they may transfer this understanding to the target system and its behavior.[17] Thus, the kind of understanding that is covered under the proposed explication of opacity is important for making scientific progress.

[17] This is presumably a reason why Evans et al. (2013) discuss the understanding of simulation outcomes in a handbook on social simulations.

Acknowledgements I'm grateful for extremely valuable comments by several anonymous referees. I'd also like to thank audiences from the "Epistemology of the Large Hadron Collider" research project, the ETH Zurich and the University of Bern. Last but not least, I'm very grateful to Christoph Baumberger for his helpful and detailed comments.

Funding Open Access funding provided by Universität Bern.

References

Alvarado, R., & Humphreys, P. (2017). Big data, thick mediation and representational opacity. *New Literary History, 48*, 729–749.

Barberousse, A., Franceschelli, S., & Imbert, C. (2009). Computer simulations as experiments. *Synthese, 169*, 557–574.

Barberousse, A., & Vorms, M. (2014). About the warrants of computer-based empirical knowledge. *Synthese, 191*(15), 3595–3620.

Baumberger, C. (2011). Types of understanding: Their nature and their relation to knowledge. *Conceptus, 40*, 67–88.

Baumberger, C. (2019). Explicating objectual understanding: Taking degrees seriously. *Journal for General Philosophy of Science, 50*(3), 367–388.

Baumberger, C., Beisbart, C., & Brun, G. (2017). What is understanding? An overview of recent debates in epistemology and philosophy of science. In S. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining understanding: New perspectives from epistemology and philosophy of science* (pp. 1–34). Routledge.

Beisbart, C. (2012). How can computer simulations produce new knowledge? *European Journal for Philosophy of Science, 2*(3), 395–434.

Beisbart, C. (2019). Should validation and verification be separated strictly? In C. Beisbart & N. J. Saam (Eds.), *Computer simulation validation. Fundamental concepts, methodological frameworks, and philosophical perspectives* (pp. 1005–1028). Cham: Springer.

Boge, F. J., & Grünke, P. (forthcoming). Computer simulations, machine learning and the laplacean demon: Opacity in the case of high energy physics. In A. Kaminski, M. Resch, & P. Gehring (Eds.), *The Science and Art of Simulation II*.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data and Society, 3*, 1–12.

Carnap, R. (1962). *Logical foundations of probability* (second edition). Chicago: University of Chicago Press. (first edition 1950).

Clark, A., & Chalmers, D. J. (1998). The extended mind. *Analysis, 58*(1), 7–19.

Collins, M., Brierley, C. M., MacVean, M., Booth, B. B. B., & Harris, G. R. (2007). The sensitivity of the rate of transient climate change to ocean physics perturbations. *Journal of Climate, 20*(10), 2315–2320.

Di Paolo, E. A., Noble, J., & Bullock, S. (2000). Simulation models as opaque thought experiments. In M. A. Bedau et al. (Eds.), *Proceedings of the 7th International Conference in Artificial Life* (pp. 497–506). Cambridge: MIT Press.

De Millo, R. A., Lipton, R. J., & Perlis, A. J. (1979). Social processes and proofs of theorems and programs. *Communications of the ACM, 22*, 271–280.

Durán, J. M., & Formanek, N. (2018). Grounds for trust: Essential epistemic opacity and computational reliabilism. *Minds and Machines, 28*(4), 645–666.

El Skaf, R., & Imbert, C. (2013). Unfolding in the empirical sciences: Experiments, thought-experiments and computer simulations. *Synthese, 190*(16), 3451–3474.

Evans, A., Heppenstall, A., & Birkin, M. (2013). Understanding simulation results. In B. Edmunds & R. Meyer (Eds.), *Simulating social complexity* (pp. 173–195). Springer.

Fernández, J. (2003). Explanation by computer simulation in cognitive science. *Minds and Machines, 13*(2), 269–284.

Gardner, M. (1970). Mathematical games—The fantastic combinations of john conway's new solitaire game "life." *Scientific American, 223*, 120–123. https://doi.org/10.1038/scientificamerican1070-120

Gordon, C., Cooper, C., Senior, C. A., Banks, H., Gregory, J. M., Johns, T. C., Mitchell, J. F. B., & Wood, R. A. (2000). The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics, 16*, 147–168. https://doi.org/10.1007/s003820050010

Gregory, J. M., & Lowe, J. A. (2000). Predictions of global and regional sea-level rise using AOGCMs with and without flux adjustment. *Geophysical Research Letters, 27*(19), 3069–3072.

Grimm, S. R. (2006). Is Understanding a species of knowledge? *British Journal of Science, 57*, 515–535.

Grimm, S. R. (2011). Understanding. In S. Bernecker & D. Pritchard (Eds.), *Routledge companion to epistemology* (pp. 84–94). Routledge.

Grüne-Yanoff, T. (2009). The explanatory potential of artificial societies. *Synthese, 169*, 539–555.

Hanappe, P., Beurivé, A., Laguzet, F., Steels, L., Bellouin, N., Boucher, O., Yamazaki, Y. H., Aina, T., & Allen, M. (2011). FAMOUS, faster: Using parallel computing techniques to accelerate the FAMOUS/HadCM3 climate model with a focus on the radiative transfer algorithm. *Geosci. Model Dev., 4*, 835–844. https://doi.org/10.5194/gmd-4-835-2011

Hills, A. (2016). Understanding Why. *Noûs, 50*, 661–688. https://doi.org/10.1111/nous.12092

Humphreys, P. (1994). Numerical experimentation. Philosophy of physics, theory structure and measurement theoryIn P. Humphreys (Ed.), *Patrick Suppes: Scientific Philosopher* (Vol. 2, pp. 103–118). Kluwer Academic Publishers.

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

Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese, 169*, 615–626.

Imbert C. (2017). Computer Simulations and Computational Models in Science. In L. Magnani & T. Bertolotti (Eds.), *Springer handbook of model-based science* (pp. 735–781), Springer, Cham. Doi: https://doi.org/10.1007/978-3-319-30526-4_34

Jebeile, J. (2018). Explaining with simulations. Why visual representations matter. *Perspectives on Science, 26*(2), 213–238.

Kaminski, A., Resch, M., & Küster, U. (2018) Mathematische Opazität. Über Rechtfertigung und Reproduzierbarkeit in der Computersimulation. In *Arbeit und Spiel* (pp. 253–278). Jahrbuch Technikphilosophie, Nomos Verlagsgesellschaft mbH & Co. KG.

Kuorikoski, J. (2011). Simulation and the sense of understanding. In P. Humphreys, & C. Imbert (Eds.), *Models, Simulations, and Representation*s. London: Routledge.

Küppers, G., & Lenhard, J. (2005). Computersimulationen: Modellierungen 2. Ordnung. *Journal for General Philosophy of Science*, *36*(2), 305–329.

Kvanvig, J. (2003). *The Value of Knowledge and the Pursuit of Understanding*. Cambridge: Cambridge University Press.

Lenhard, J. (2006). Surprised by a nanowire: Simulation, control, and understanding. *Philosophy of Science, 73*(5), 605–616.

Lenhard, J. (2009). The Great Deluge. Simulation modeling and scientific understanding. In H. W. de Regt, S. Leonelli, & K. Eigner (Eds.), *Scientific Understanding. Philosophical Perspectives* (pp. 169–186). Pittsburgh: University of Pittsburgh Press.

Lenhard, J. (2011). Epistemologie der Iteration. Gedankenexperimente und Simulationsexperimente. *Deutsche Zeitschrift für Philosophie*, *59*(1), 131–145.

Lenhard, J. (2019). *Calculated Surprises*. Oxford University Press.

Newman, J. (2016). Epistemic opacity, confirmation holism and technical debt: Computer simulation in the light of empirical software engineering. In Gadducci, F. & Tavosanis, M. (Eds.), *History and Philosophy of Computing. HaPoC 2015.* IFIP Advances in Information and Communication Technology, vol. 487 (pp. 256–272). Cham: Springer.

Oberkampf, W. L. (2019). Simulation Accuracy, Uncertainty, and Predictive Capability: A Physical Sciences Perspective. In C. Beisbart, & N. J. Saam (Eds.), *Computer Simulation Validation. Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives* (pp. 69–97). Cham: Springer.

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science, 263*, 641–646.

Parker, W. (2010). Whose probabilities? Predicting climate change with ensembles of models. *Philosophy of Science, 77*, 985–999.

Parker, W. S. (2014). Simulation and understanding in the study of weather and climate. *Perspectives on Science, 22*(3), 336–356.

Pope, V. D., Gallani, M. L., Rowntree, P. R., & Stratton, R. A. (2000). The impact of new physical parametrizations in the Hadley Centre climate model—HadAM3. *Climate Dynamics, 16*, 123–146. https://doi.org/10.1007/s003820050009

Pritchard, D. (2010). Knowledge, understanding and epistemic value. In A. O'Hear (Ed.), *Epistemology* (pp. 19–43). Cambridge University Press.

Quine, W. V. O. (1953). Reference and Modality. In W. V. O. Quine (Ed.), *From a Logical Point of View* (pp. 139–157). Cambridge (MA), here quoted from the revised edition 1980.

Rider, W. J. (2019). The Foundations of Verification in Modeling and Simulation. In C. Beisbart, & N. J. Saam (Eds.), *Computer Simulation Validation. Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives* (pp. 271–293). Cham: Springer.

Roache, P. (1997). Quantification of uncertainty in computational fluid dynamics. *Annual Review of Fluid Mechanics, 29*, 123–160.

Roache, P. R. (2019). The Method of Manufactured Solutions for Code Verification. In C. Beisbart, & N. J. Saam (Eds.), *Computer Simulation Validation. Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives* (pp. 295–318). Cham: Springer.

Saam, N. J. (2017). Understanding social science simulations: Distinguishing two categories of simulations. In M. Resch, A. Kaminski, & P. Gehring (Eds.), *The science and art of simulation I. Exploring—Understanding—Knowing* (pp. 67–84). Cham: Springer.

San Pedro, I. (forthcoming). Degrees of epistemic opacity. In M. Resch, A. Kaminski, & P. Gehring (Eds.), *Epistemic opacity in computer simulation and machine learning*.

Schlesinger, S., et al. (1979). Terminology for model credibility. *SIMULATION, 32*, 103–104.

Symons, J., & Alvarado, R. (2016). Can we trust big data? Applying philosophy of science to software. *Big Data and Society, 3*(2), 1–17.

Teller, P. (1980). Computer proof. *The Journal of Philosophy, 77*(12), 797–803.

Turkle, S. (1997). Seeing through computers. *The American Prospect, 8*(31), 76–82.

Turkle, S. (2004). How computers change the way we think. *The Chronicle of Higher Education, 50*(21), B26–B28.

Tymoczko, T. (1979). The four-color problem and its philosophical significance. *The Journal of Philosophy, 76*(2), 57–83.

Wagenknecht, S. (2014). Opaque and translucent epistemic dependence in collaborative scientific practice. *Episteme, 11*(4), 475–492.

Walton, K. L. (1984). Transparent pictures: On the nature of photographic realism. *Noûs, 18*(1), 67–72.

Winsberg, E. (1999). Sanctioning models. The epistemology of simulation. *Science in Context, 12*, 275–292.

Wolfram, S. (2002). *A new kind of science*. Wolfram Media Inc.

Ylikoski, P. (2014). Agent-based simulation and sociological understanding. *Perspectives on Science, 22*(3), 318–335.