




Why computer simulations are not inferences, and in what sense they are experiments

Florian J. Boge^{1,2} 

Received: 15 March 2018 / Accepted: 11 November 2018 / Published online: 29 November 2018

© The Author(s) 2018, corrected publication 2022

Abstract

The question of where, between theory and experiment, computer simulations (CSs) locate on the methodological map is one of the central questions in the epistemology of simulation (cf. Saam *Journal for General Philosophy of Science*, 48, 293–309, 2017). The two extremes on the map have them either be a kind of *experiment* in their own right (e.g. Barberousse et al. *Synthese*, 169, 557–574, 2009; Morgan 2002, 2003, *Journal of Economic Methodology*, 12(2), 317–329, 2005; Morrison *Philosophical Studies*, 143, 33–57, 2009; Morrison 2015; Massimi and Bhimji *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 51, 71–81, 2015; Parker *Synthese*, 169, 483–496, 2009) or just an *argument* executed with the aid of a computer (e.g. Beisbart *European Journal for Philosophy of Science*, 2, 395–434, 2012; Beisbart and Norton *International Studies in the Philosophy of Science*, 26, 403–422, 2012). There exist multiple versions of the first kind of position, whereas the latter is rather unified. I will argue that, while many claims about the ‘experimental’ status of CSs seem unjustified, there is a variant of the first position that seems preferable. In particular I will argue that while CSs respect the *logic* of (deductively valid) arguments, they neither agree with their *pragmatics* nor their *epistemology*. I will then lay out in what sense CSs can fruitfully be seen as experiments, and what features set them apart from traditional experiments nonetheless. I conclude that they should be seen as *surrogate experiments*, i.e. experiments executed consciously on the wrong kind of system, but with an exploitable connection to the system of interest. Finally, I contrast my view with that of Beisbart (*European Journal for Philosophy of Science*, 8, 171–204, 2018), according to which CSs are surrogates *for* experiments, arguing that this introduces an arbitrary split between CSs and other kinds of simulations.

Keywords Epistemology of simulation · Argument view · Causal interaction claim

✉ Florian J. Boge
fboge@uni-wuppertal.de

1 Introduction

1.1 State of play

Computer simulations (CSs) play an integral role in modern science. They are used, e.g., to model the atmosphere in climate science, investigate stability properties of cars, buildings, and other pieces of engineering, and are involved in the design, execution, and evaluation of highly complex experiments in high energy physics (HEP). A recent review (Saam 2017, cf. p. 295) displays the question of where between theory and experiment CSs locate on the methodological map¹ as one of the central questions in the epistemology of simulation. The two extremes on the map have them either be a kind of *experiment* in their own right (e.g. Barberousse et al. 2009; Morgan 2002, 2003, 2005; Morrison 2009, 2015; Massimi and Bhimji 2015; Parker 2009), possibly with an epistemological status comparable to that of a traditional laboratory experiment, or just an *argument* executed by (or with the aid of) the computer (cf. Beisbart 2012; Beisbart and Norton 2012).

Whereas the latter position is rather unified, there exist multiple versions of the first one. I here want to focus on two such versions: the one which I find to be the boldest (Morrison 2015; Massimi and Bhimji 2015) and one which I find to be among the most modest ones (Parker 2009; Winsberg 2009; Dardashti et al. 2015). I will argue that (i) while CSs can quite generally be *reconstructed* as arguments and sometimes at the same time *replace* experiments, this neither sanctions that their *epistemic power* is that of an argument or an experiment respectively; and (ii) that CSs can be profitably viewed as *surrogate experiments*, making them epistemically inferior to laboratory experimentation but practically often preferable.

The paper is organized as follows: Section 2 exposes the argument view and the ‘boldest’ experiment view in detail (Sections 2.1 and 2.2). It is then (Section 2.3) demonstrated that central arguments in favor of each view are wanting, because evidence advanced in support of the central epistemological hypothesis of each camp does not yield the desired support. Section 3 consecutively establishes which insights can and cannot be maintained from both camps (Sections 3.1 and 3.2) in the light of the foregoing discussion. Finally Section 4 establishes a modest experiment view (Section 4.1) and defends this view against a recently proposed alternative (Section 4.2). To further the discussion, I will also give a brief analysis of the basic ingredients to any CS below.

1.2 Anatomy of a CS

A CS, in short, comprises the execution of a set of ‘rules’, the algorithms contained in some code, by a physical system, the programmable digital computer. This execution may be called a *simulation step*, resulting in an *output* (an expectation value, an array of numbers, some visualization in the form of a graph, an animation, a color

¹Saam borrows this term from Galison (1996, p. 120); cf. her p. 293.

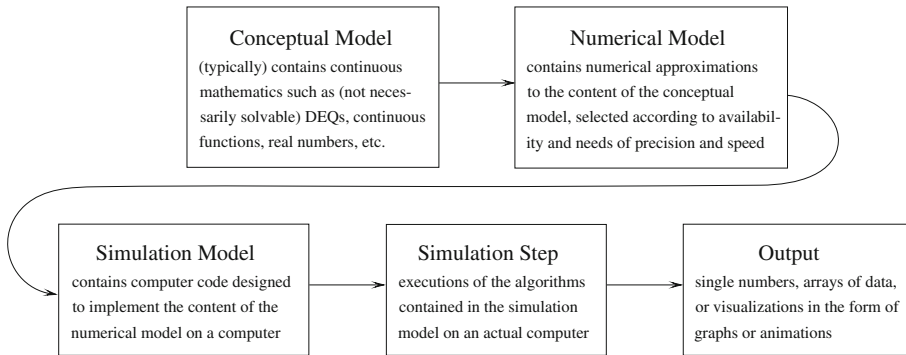


Fig. 1 Dissection of the process of devising a CS into different logical and performative components

plot...), the set of algorithms a *simulation model*.² This simulation model will usually (if not always) be based on some previously existing *numerical*, i.e., *discrete mathematical model* of a system of interest (the ‘target system’), which in many cases is an approximation to another model based on *continuous* mathematics, and hence not suited for a translation into algorithms.

The first model in this chain I call, following Morrison (2015, p. 254), a *conceptual model*. It may or may not coincide with the numerical one, depending on the subject matter.³ The dependencies between these modeling and other steps are illustrated in Fig. 1.⁴ This should not be over-interpreted as depicting any actual temporal sequence: The process of devising a CS is typically not as linear as depicted in Fig. 1, but will contain multiple loops. For instance: once the output has some recognizable flaws, there are multiple junctions at which revisions are possible. And it is equally possible that parts of a code will be already implemented and executed before the entire simulation model is complete, e.g. to test features of these very parts or to fit free parameters contained therein. Figure 1, in other words, depicts an ideal limit in which one is entirely certain as to what to simulate and how.

Notably, the diagram exhibits some cross-categorial features: the models involved are ‘purely logical’, the simulation step is performative. Both these aspects are relevant to a proper understanding of CSs from an epistemological point of view, as the discussion will show.

²Morrison (2009, 2015) uses this term for what I call a numerical model below and calls what I have called a simulation model a numerical algorithm. This terminology seems unfortunate, to the extent that the algorithm(s) implemented ultimately constitute a *different* – albeit connected and, in some sense to be spelled out, derivative – *model*. The widely acknowledged fact that the content of a given numerical model “has to be transformed (sometimes *drastically*) into algorithms” (Barberousse et al. 2009, p. 558; *emph. added*) should provide enough credence for the existence of a distinct such simulation model.

³In physics, for instance, a conceptual model will typically derive from an underlying theory, formulated with heavy reliance on differential equations and integrals. Schelling’s model of agent segregation (e.g. Winsberg 2015, sect. 2.2) or models underlying simulations of population dynamics in biology, in which the evolution proceeds in discrete time steps (generations), are examples in which the conceptual model is already ‘numerical’ or discrete.

⁴A similar illustration, that is however discernibly different in important respects, can be found in Winsberg (2010, p. 11).

2 CSs as arguments vs. CSs as experiments

2.1 The argument view

A particularly interesting view of CSs with an inherent initial plausibility is the view of Beisbart (2012) and Beisbart and Norton (2012) of CSs as *arguments*.

For CSs implementing a deterministic evolution, the case is straightforward. Consider some differential equation (DEQ) and a function found to solve it, given suitable boundary conditions. Solving the DEQ by that function involves only deductively valid steps (mathematical computations), and since the function solving the DEQ will be used to describe the behavior of some system (the trajectory of some particle, say), this can be easily viewed as, or translated into an argument from the assumption of certain dynamics and prevailing conditions (the argument's premises) to an allowed description of the behavior of the system in question (the conclusion).⁵

There are two obvious subtleties involved: (a) neither the numerical nor the simulation model is strictly identical with the conceptual one, so long as the latter is continuous; and (b) the statement that any of the models, conceptual, numerical, or simulation, 'can be easily viewed as, or translated into' an argument clearly needs substantiation. That "[e]ach computer simulation can be reconstructed as an argument" is called the *reconstruction thesis* by Beisbart (2012, p. 403). As formulated by him, however, there is a second part to it, namely that "the epistemic power of the CS is that of the argument." (ibid.) These are clearly two distinct claims. I will here only think of the first part as 'the reconstruction thesis', and address the second part, which I will call the *epistemic power thesis*_A (short: EPT_A), separately.

Moreover, I will here understand 'argument' as referring to a deductively valid inference: one in which the truth of some statement, the conclusion, can be inferred *with certainty* from the truth of a set of (other) statements, the premises. This is the notion of 'argument' at stake with the accounts of Beisbart (2012) and Beisbart and Norton (2012).⁶

To substantiate the reconstruction thesis in the deterministic context, Beisbart (2012, p. 405 ff.) provides an example wherein the DEQ for a damped and driven pendulum is approximated by the Euler method. The initial conditions are stated as a conjunctive premise; the discrete time steps as conditional ones, connecting the evolution at some given time step (antecedent) with the one at the subsequent time step (consequent). The conclusion then is a conjunction of statements describing the concrete values computed for these times. This is just an example, but Beisbart (2012, p. 414 ff.) also demonstrates how algorithms can be translated into arguments more generally.

⁵I have been careful to talk merely of the *allowed* behavior of the system since, as is well known, for many DEQs there exist no secure methods for finding *unique* solutions, and in some cases there even provably are none (e.g. Nagle et al. 2012, p. 15 for an example). The corresponding mathematical argument hence often establishes only the *existence* of a particular trajectory or so, and only sub specie aeternitatis (no system may ever exemplify that particular trajectory).

⁶There are multiple mentions of *inductive* arguments in Beisbart and Norton (2012) as well, but from the way the term is used I find that there may be a confusion at play, as I shall lay out in detail later.

As for subtlety (a), Beisbart (2012, p. 406) and Beisbart and Norton (2012, p. 410) refer back to theorems securing convergence behavior or quantifying errors implied by a technique. Such theorems allow to infer similarity or quantitative proximity between continuous and numerical models for a given technique at a suitable grain. They hence correspond to *meta-arguments* (cf. Beisbart and Norton, 2012, p. 415) that establish a connection between the result of a numerical model (or the conclusion of the corresponding ‘numerical argument’) and the result of a conceptual model (the conclusion of the argument originally intended). Given that one is interested in a certain precision only, one can then use such meta-arguments to ‘settle for’ the conclusion derived from the numerical argument.

Monte Carlo simulations (MCSs) prima facie defy the reconstruction thesis. A Monte Carlo (MC) technique, to recall, “is any technique making use of random numbers⁷ to solve a problem.” (James 1980, p. 1147) Because of this involvement of randomness, MCSs are sometimes referred to as “almost experimental” (Krauth 2006, p. 1), and one might think that they cannot possibly be reconstructed as deductively valid arguments.

This is a false impression. First note that different MC techniques (e.g. Thijsen 2007, p. 285, for an overview) all ‘boil down’, in a sense, to the computation of an MC *integral* (cf. James 1980, p. 1148, for details), meaning the sum of a function f over some interval $[a, b]$, in the simplest case over N values of f for randomly chosen evaluation points, multiplied by $(b - a)/N$. In probabilistic models, moreover, the values of all interesting, comparable quantities will be given as averages $\bar{A} = \frac{1}{N} \sum_{i=1}^N A_i$, with N the number of steps in the corresponding MC algorithm (cf. Thijsen 2007, pp. 192 and 302, for an example); formally an MC integral with a convenient choice of variables.

This correspondence between MC techniques and integrals is exploited by Beisbart and Norton (2012, p. 413 ff.) to reconstruct the simulation of a Brownian particle in terms of arguments: They first reconstruct the simulation of a set of trajectories of the particle by a set of arguments like the one for the Euler method, with the successive steps determined by the random numbers of a MC algorithm. Then a subsequent argument provides the expected final position of the particle after N steps by appeal to the set of single-trajectory conclusions of the former arguments.

It is inessential that the steps are chosen at random or with no recognizable pattern: it is the deductive validity of an inference from initial to final position given the steps so selected that makes reconstruction in terms of an argument possible. More precisely, while the steps are *chosen* at random, they *are* by no means random; all possible steps are laid out in the prescriptions defining the algorithm, which in turn depends on the probability model of Brownian motion.

⁷De facto *pseudo random* numbers are often used, i.e. sequences of numbers generated by some algorithm that exhibit a discernible *non-random pattern* only after a suitably long period (e.g. Thijsen 2007, p. 605 ff.). The factual non-randomness is a difficulty that equally needs to be tackled by appeal to meta-arguments. Cf. Beisbart and Norton (cf. 2012, p. 407) and James (1980, p. 1169 ff.) for details.

2.2 The experiment view at its boldest

As I have detailed in the introduction, there is not one unified view of CSs as experiments, but rather a plethora of different views that fit under that label. I here want to focus on the ‘boldest’ view of this kind, namely of CSs as experiments that are *epistemically on a par* with traditional laboratory experiments in certain circumstances. Now if CSs can themselves be viewed as an experimental activity, and if that experimental activity can be epistemically on a par with traditional laboratory experimentation, then a CS can clearly replace a laboratory experiment under favorable circumstances. As with the reconstruction thesis, I see two aspects here that need to be carefully distinguished: (i) that CSs can de facto *function* as replacements of laboratory experiments in scientific research, and (ii) that this replacement goes *without epistemic loss*. I will refer to (i) as *the replacement thesis*, and to (ii) as the *epistemic power thesis_E* (short: EPT_E), as it claims that the epistemic power of a CS (of the right kind) is that of the experiment replaced.

The replacement thesis is at the heart of arguments given by Morrison (2015, p. 241), who addresses the “issue [...] whether simulation data can replace experimental measurement in certain contexts” in detail, or Massimi and Bhimji (2015, p. 79), who take “scientists’ choice as to whether to use simulations or experiments” for certain purposes in certain contexts to demonstrate “the interchangeable role they play” therein. That this thesis is somewhat plausible in itself can also be illustrated quite nicely on a toy example: Imagine an engineer designing a new building. Having devised the basic construction scheme, she will certainly want to investigate the stability properties of her construction without having to build multiple versions and tear them down again. In such cases, CSs will be used to gather the required data, and will hence replace actual experiments.

Moreover, based on their investigation of the Higgs-discovery by CERN’s ATLAS experiment, Massimi and Bhimji (2015) conclude that “computer simulations are on a par with / interchangeably used with experiments at ATLAS” (p. 81) in specified contexts. This outright identification gives credence to the fact that evidence for the replacement thesis is taken to be direct evidence for EPT_E .

The origin of EPT_E is probably Morrison (2009, p. 55-6), where she argues that “simulation can attain an epistemic status comparable to laboratory experimentation”, and at the same time seeks “justification for classifying simulations *as* experiments” (ibid., p. 54; *emph. added*). Similar views are defended by Massimi and Bhimji (2015), who investigate “the extent to which [CSs – FJB] count as epistemologically on a par with traditional experiments” (p. 71), and in the next sentence point out that “[c]ritics have raised doubts about simulations *being* genuine experiments[...].” (ibid.; *emph. added*) This gives credence to the fact that EPT_E is entertained, in the relevant literature, as a kind of view of CSs *as* experiments.

The identification is, of course, not unmotivated: Morrison (2009) describes in detail how the measurement of the gravitational acceleration, g , requires an elaborate *model* of the entire setup, and how g is actually *calculated* from cord length ℓ and period of oscillation T .

So the ‘measurement’ of g comprises modeling, performative, and calculational steps, not unlike a CS in my reconstruction from Section 1.2. From this similarity

and the involvement of elaborate models in both cases derives the point Morrison (2009, p. 52) wants to make: “Given this [...] role of models in experiment, what, if anything, differentiates experiment from simulation?” And there is a further commonality between CSs and (laboratory) experiments, namely, *causal contact* with the target system:

The programme itself is tested and calibrated to ensure that it reproduces the behaviour of the target system [...]. As in an experiment we can trace the linkages back to some material features of the target system that are modelled in the initial stages of investigation. (Morrison 2009, p. 53)

Thus both experiments and CSs are also causally linked to the target system. If all this is the case, it becomes understandable how our earlier imagined engineer might rely on CSs to investigate the static properties of her prospective building. Should they then, if they are so much like measurements or experiments, not rather count as a subspecies of the latter?

2.3 Replacement and reconstruction in action, simultaneously

To make their point, both Massimi and Bhimji (2015) and Morrison (2015, ch. 8) focus on a case study from high energy physics (HEP), namely the discovery of the Higgs boson at CERN’s Large Hadron Collider (LHC). The reason is that CSs are involved in virtually every step of experiments at the LHC, not least due to the complexity of the measured signals and statistical analyses.

Such statistical analyses in LHC physics will inevitably involve taking into account contributions from the detector. A central quantity here is the so called *transfer function* $\mathcal{T}(x|y)$, describing “the probability to observe x [...] given that the true value was y .” (Cowan 1998, p. 156) y could be, say, the energy \tilde{E}_{e^-} of an electron resulting from the leptonic decay of some intermediate vector boson, and x would then be the energy E_{e^-} measured in the detector.

Transfer functions may include limited efficiency contributions due to detector geometry (i.e. angles without detection volume), known errors, or lower energy thresholds for reaction, all usually accumulated into an *efficiency function* $\varepsilon(y)$. ε determines the probability that value y will lead to any measurement in the first place. The remainder of $\mathcal{T}(x|y)$ will then be called a *resolution function* $s(x|y)$ (cf. Cowan 1998, *ibid.*).

Now it is in fact a standard practice to determine $\mathcal{T}(x|y)$ *not* by calibration of the *real* detector, but “by using a Monte Carlo simulation based on an understanding of the physical processes that take place in the detector.” (Cowan 1998, p. 157) This means a determination of unknown (‘bulk’) properties of the detector based on one’s background knowledge of the latter. The situation is hence very much like that of the engineer and her building we had imagined earlier: the physics of the detector is assumed to be so well known that a simulation suffices to figure out the relevant missing detail; no additional laboratory measurements are performed or assumed to be necessary. The replacement thesis applies perfectly well.

The crucial thing to realize is that there is also no problem with reconstructing this simulation in terms of an argument. An MC event generator will produce ‘data’

whose occurrence can be reconstructed by premises of the form ‘The energy of the e^- is \hat{E}_{e^-} ’. This data will be subjected to a chain of further simulations modeling the responses of parts of the detector.⁸ As mentioned before, any such simulation will already contain probabilistic information about the different *parts*, which information could again be expressed in the form of resolution functions and efficiencies (i.e. partial transfer functions).

Hence $\mathcal{T}(x|y)$ will describe a *cumulative* effect of all these parts acting in concert, and the background knowledge about different parts of the detector could be provided in terms of pairs $p_i = (\varepsilon_i, s_i)$ of efficiencies and resolution functions, each defining a *probability model* for part i respectively.

Neglecting, for simplicity, the aspect of efficiency that is given by pure detector geometry, I assume that the effect of the efficiencies ε_i can be modeled entirely by a hit-or-miss technique. This means that on any run the i th part of the detector will respond to the energy of the electron if and only if the value for ε_i at the energy the electron has before it enters part i exceeds a certain number $g \in [0, m)$ that is chosen by a uniform random number generator, and where m is the maximum value that ε_i assumes (e.g. Lista 2017, pp. 76-8).

According to these considerations, the argument reconstructing the determination of $\mathcal{T}(x|y)$ follows quite closely the pattern of that reconstructing the determination of the Brownian particle’s expected position as provided by Beisbart and Norton (2012, pp. 414-5). For any value of \hat{E}_{e^-} considered, there will be a set of N arguments of the following form, where i ranges over the D parts of the detector contributing efficiency and resolution effects:

- P1 The initial energy of e^- is $\hat{E}_{e^-} \equiv E_{e^-}^0$.
- P2(i) e^- interacts with part i of the detector.
- P3(i) If e^- interacts with part i of the detector, this part will respond to the interaction if and only if $\varepsilon_i(E_{e^-}^{i-1}) > g_i$, where $g_i \in [0, m)$ samples a uniform distribution, and m is the maximum value that ε_i assumes.
- P4(i) $\varepsilon_i(E_{e^-}^{i-1})$ is / is not greater than g_i .
- P5(i) If part i of the detector responds to the interaction, the energy of e^- will be transmitted as $E_{e^-}^i = \hat{E}_{e^-} + \sum_{j=1}^{i-1} \delta_j + \delta_i$, where δ_i samples s_i .
- P6 If one part of the detector does not respond, the detector will not provide an output.
- C The energy of e^- will be transmitted as $E_{e^-}^D \equiv E_{e^-}$. / The detector will not provide an output.

⁸I focus on ‘parts’ (meaning specific components like calorimeter cells etc.) because in some cases sufficiently detailed information about the microphysics is unavailable (e.g. due to unknown details of the composition) and one has to resort to bulk probabilistic information from measurements (cf. GEANT 4 for a rich source of details). Otherwise one could use computations from the models (phenomenologically) describing the microphysics to define a probability model, and reconstruct the determination of partial transfer functions in more or less the same way as is done for the overall one below.

To find the respective probabilities, there is the need to appeal to a subsequent argument at any value of \hat{E}_{e^-} – drawing on the N arguments of the above form. For N large enough, these subsequent arguments can be assumed to be suitably backed up by probabilistic convergence, and will be of the form:

- P1 The energy of e^- is \hat{E}_{e^-} and n_1 out of N times this will lead to the measured value $E_{e^-}^1$, n_2 out of N times to $E_{e^-}^2$, ..., n_ℓ out of N times to $E_{e^-}^\ell$, ..., and n_K out of N times to $E_{e^-}^K$ ($N = \sum_{\ell=1}^K n_\ell$).
- P2(ℓ) If the energy of e^- is \hat{E}_{e^-} and this will lead to the measured value $E_{e^-}^\ell$ n_ℓ out of N times, then $\mathcal{T}(E_{e^-}^\ell | \hat{E}_{e^-})$ is very probably approximately n_ℓ/N .
- C The probabilities $\{\mathcal{T}(E_{e^-}^\ell | \hat{E}_{e^-})\}_{1 \leq \ell \leq K}$ for the measured values $\{E_{e^-}^\ell\}_{1 \leq \ell \leq K}$ at \hat{E}_{e^-} are very probably approximately $\{\frac{n_\ell}{N}\}_{1 \leq \ell \leq K}$.

As should be obvious, any such CS must select a discrete set of values \hat{E}_{e^-} and a finite number of runs N , as well as a finite number of alterations δ_i to the energy for each part. In case one is interested in a continuous $\mathcal{T}(E_{e^-} | \hat{E}_{e^-})$ at any \hat{E}_{e^-} one would hence have to interpolate. This interpolation could again be reconstructed by an argument that appeals to premises of the rough form ‘If the probabilities at \hat{E}_{e^-} are very probably approximately $\{\frac{n_\ell}{N}\}_{1 \leq \ell \leq k}$, then very probably the transfer function $\mathcal{T}(E_{e^-} | \hat{E}_{e^-})$ is given by ...’⁹

Such reconstructions are a somewhat tedious but straightforward exercise. The one just given establishes not only that reconstruction is easily possible (on a somewhat coarse level) in more involved cases than Brownian particles, but also that both theses, the reconstruction thesis and the replacement thesis, can fit perfectly well at the same time: There is a reconstructing argument for a case in which a CS is used in the place of an experiment; and there is no reason why reconstruction should not be generally possible in such cases.

That this co-occurrence is possible means that both theses are merely *complementary*, not mutually contradictory. But EPT_A and EPT_E are, so it is vital to realize that the reconstruction and replacement theses are only *necessary* conditions for EPT_A and EPT_E respectively, not sufficient ones: *If* the epistemic power of a CS is just that of a particular argument, then *reconstruction* of the CS by that argument is clearly possible; *if* its epistemic power is that of a corresponding experiment, then *replacement* is clearly possible. It is unclear that either of the *converse* claims holds.

Bottom line: It is important to distinguish EPT_A and EPT_E from the reconstruction and replacement theses respectively, and evidence for the latter two cannot provide *direct* evidence for the former two.

⁹In fact, this would provide a probability distribution for E_{e^-} given \hat{E}_{e^-} . In case the transfer function is specified as a probability *density*, as it standardly is, it would be the derivative of that distribution.

3 What to gain from both camps

3.1 Putting replacement in its place

I have given reasons to doubt that the evidenced possibility of replacing certain lab experiments by CSs sanctions the inference to CSs being epistemically on a par with those experiments (to EPT_E). How does EPT_E fare independently of its support by the replacement thesis? Now there is an obvious asymmetry between the replacement and the reconstruction thesis, which, in fact, points to limitations of EPT_E . For while it is hard to think of examples where a CS cannot in principle be reconstructed as an argument, could we have equally measured, say, the Higgs mass by means of CSs alone, just as we can thereby ‘measure’ the static properties of buildings, the responses of detectors, or expected backgrounds in HEP (cf. Massimi and Bhimji 2015, p. 77)? Given that the implication is from EPT_E to the replacement thesis, not vice versa, the failure of being able to replace *all* lab experiments by CSs implies at least limitations to EPT_E .

Recall, however, that Morrison’s claim merely was that “simulation *can* attain an epistemic status *comparable* to laboratory experimentation[...]” (emphasis added) So the claim never was to an unrestricted validity of EPT_E in the first place, but rather to its applying in favorable cases. EPT_E may still be valid, for instance, in the situation of the imagined engineer: Given the *well-known* laws of statics, she can find out a lot about her building without having to experiment on it directly. The point is that replacements can occur when one has sufficient inductive support for one’s simulation model or the models from which it derives.

What is really wrong with viewing CSs as epistemically on a par with laboratory experimentation (and field observation) can be brought out as follows. Assume that the laws of nature would suddenly change, or that we had some of them wrong in the first place, in such a way that a clever series of contrived laboratory experiments would bring out the difference to accepted dogma. No CS could ever bring this change to the light by itself. Only after having understood the alterations to the laws from traditional experiments could CSs based on these newly found laws be used to increase our knowledge further.¹⁰ It is in this sense that laboratory experimentation enjoys an epistemic priority over CSs: CSs have to ‘answer to’ lab experiments in way which does not go the other way around; and lab experiments have an in-principle ability to extend our knowledge in way that CSs cannot.

This sort of reasoning in fact connects to two well-established (and equally connected) lines of argument in the literature. The first one is what has been called the *causal interaction claim* (CIC) by Massimi and Bhimji (cf. 2015, p. 74), first advanced by Giere (2009) against Morrison (2009):

I just do not see how this similarity between traditional experiments and computer experiments puts them epistemologically on a par. The epistemological payoff of a traditional experiment, *because of the causal connection with the*

¹⁰Put this way, it seems insufficient to have “enough of a toolkit of trustworthy model building principles” (Winsberg 2009, p. 591) to establish the epistemic equality of CSs with lab experiments.

target system, is greater (or less) confidence in the fit between a model and a target system. A computer experiment, which does not go beyond the simulation system,¹¹ has no such payoff. (Giere 2009, p. 61; *emph. added*)

Morrison's reasoning has been defended in detail against the CIC by Massimi and Bhimji (2015, p. 74; *orig. emph.*), who present three different versions of it:

(CIC₁) Experiments involve *direct* causal interactions with the target system when a physical quantity is *calibrated* by direct comparison with observed data.[...]

(CIC₂) Experiments involve *quasi-direct* causal interactions with the target system when the experimental apparatus is designed to *track* how a physical quantity may interact with another, suitably chosen.[...]

(CIC₃) Experiments involve *indirect* causal interactions with the target system when we *infer* an entity against relevant experimental background.

They then (p. 80) argue that CIC₁ and CIC₂ apply to simulations all the same and that due to the heavy de facto dependence of e.g. the Higgs detection on simulations of the expected background and in the interpretation of the data, “[t]he thesis of the epistemological priority of ordinary experiments over computer simulations [...] seems to lose its bite also along CIC₃ lines.” (p. 81)

Now I have reservations about the claims concerning CIC₂ and CIC₃, in the latter case because many physicists believe that the Higgs $\rightarrow \gamma\gamma$ -channel could have in principle been used to find the Higgs without any involvement of CSs at all,¹² and in the former case because the ‘tracking’ appealed to by Massimi and Bhimji in the case of CSs is entirely internal to the simulation model, whereas the claims to measurement or discovery) come about only by comparison with experimental data. But the more fundamental flaw, in my opinion, is that none of CIC_{1–3} even makes contact with the real bite of Giere's argument.

The crucial point of Giere's allusions to causal contact with the target system actually connects directly to the second line of argument mentioned above. It is that in a laboratory experiment involving the target system itself, *it* can ‘bite back’ in an unexpected way. This is a version of Morgan's (2003) *confoundment argument*:

in the laboratory, there is always the possibility of not only being surprised but of being confounded, for the world in the laboratory is one where not only are we ignorant of the outcomes, we don't even know in advance everything about the behavior of the material elements being used.¹³

In a corresponding CS there is only a ‘virtual’ target system, which can only bite back in accordance with the parameters used to design it in the first place. Any

¹¹By simulation system Giere, following Morrison (2009, p. 55), means “a computer, an operating system, and an application program in which the code [...] is written.” (Giere 2009, p. 60) While I have not laid focus on the operating system or the application program, these are certainly preconditions for the implementation of some simulation model.

¹²This is at least more or less acknowledged by Massimi and Bhimji (2015); cf. their p. 81.

¹³I eschew a discussion of the allusions to ‘materiality’, as I find that they have done more harm than good to the debate.

unexpected ‘recalcitrance’ occurring in a CS must therefore either be attributed to unrecognized (maybe unintended) elements of the models used, or to unrecognized properties of the computer itself (failures of the hardware or the operating system, limitations of the programming language...). It *cannot* be contributed by the target system. Thus even though CSs can make causal contact with their target systems through calibration,¹⁴ this is not the right *kind of* causal contact to establish epistemic equality.¹⁵

It should be noted that I have thus only established reasons to reject the notion that CSs are epistemically on a par with traditional (laboratory or field) experiments (EPT_E). This does not imply that there is no sense in which they can themselves be seen as some kind of experiment. I will turn to this option in detail in Section 4.1.

3.2 Putting reconstruction in its place

Like EPT_E , EPT_A is not really sanctioned by the evidenced possibility of reconstructions. But as with EPT_E one should assess EPT_A 's independent plausibility as well. Note first that an apparent tension arises from the dynamical or ‘performative’ character of the simulation step and the static, ‘logical’ character of the models involved (cf. Section 1.2). This tension is addressed by Beisbart with what he calls the “practice thesis”: “If a computer simulation is run, the reconstructing argument is executed.” (Beisbart 2012, p. 403)

Beisbart, however, acknowledges a difficulty here, since when “we run through an argument or [...] reason, we consciously move from the premises to the conclusion”, while “nothing like this happens, when a scientist runs a computer simulation.” (Beisbart 2012, p. 420) In particular, executions of arguments are basically sequences of thoughts while “it is not always realistic to assume that the information in the memory [of the computer – FJB] is always updated sequentially. This assumption is obviously false for parallel computing.” (Beisbart 2012, p. 422)

To establish his practice thesis in spite of this, Beisbart (2012, pp. 420 ff.) appeals to the *extended mind hypothesis*, which basically says that instruments involved in cognitive tasks can be viewed as *part* of the cognizing system. Following Wedgwood (2006), Beisbart holds that reasoning is a causal process which separates into basic steps, and assuming the extended mind hypothesis, these steps can be executed by the joint agent-computer system in a way that need not conform to the conscious experience of going through an argument.

The extended mind hypothesis, however, has been met with serious criticisms, for instance that it implies a loss of first person authority over one’s own beliefs.¹⁶

¹⁴Actually, CSs are usually not even *calibrated* on processes directly involving the target system (e.g. the Higgs itself); so the notion of ‘target system’ must be meant quite loosely (elementary particles in general, or even colliders and detectors in the presence of such) in allusions to calibration as a suitable sort of causal contact.

¹⁵Parker(2010, pp. 42-3) similarly doubts that the sort of causal contact may be the right one, but for different reasons. She also somewhat resolves her prima facie reasons for not thinking of CSs as measurements in her 2017 (cf. p. 281 ff.).

¹⁶Cf. Menary (2010, p. 9 ff.) for a detailed exposition of this and further criticisms.

Moreover, Beisbart's claim was that the *computer* executes the argument, while appeal to the extended mind hypothesis implies that the *joint* agent-computer system is involved. For these reasons, appeal to the extended mind hypothesis is rather unfortunate.

This all seems quite unnecessary for understanding how there can be a reconstructing argument associated with a CS: the brain states somehow connected with the conscious execution of an argument may themselves be quite unlike the conscious experience of going through the argument.¹⁷ This usually does not interest us much so long as by 'argument' we merely mean 'deductively valid inference'. We are then typically interested in the relation of *truth-preservation* between premises and conclusion.

These considerations allow us to delineate in what sense we may call the simulation step of a CS the 'execution of an argument by the computer', regardless of the precise physical realization of this 'execution' and similar such details. It constitutes a process of *some* sort that is capable of connecting a set of statements with another statement in such a way that the connection transfers the truth of the initial set to the final statement.

At the same time it thus seems doubtful that we should take this 'execution of an argument' all too literal. To see this more clearly, consider Beisbart's (2012, p. 401) assessment of the potential novelty of a CS's outcome. Here he has it that "the assumption that knowledge is closed [under deduction] is not very plausible [...], and when we run through a sound argument, the conclusion can at least be new in the sense that we did not believe it before." This, Beisbart (2012, *ibid.*) thinks, leads to a "psychological sense of novelty[...]"

Now in providing an argument to some conclusion, we are already convinced of the conclusion's truth. For while there may be explorative phases in mathematics or logic where one 'plays around' with the deductive consequences of one's definitions and axioms, in no serious proof of a theorem would the conclusion come unexpectedly; one *first* has a mathematical conjecture and *then* tries to prove it from one's definitions and axioms. Hence as soon as the term 'argument' is taken seriously, the appeal to closure of knowledge under logical entailment seems misplaced in this context.

The same of course holds for arguments concerning beliefs with empirical content: One is lead, on the basis of one's however selective evidence, to the belief in some statement and then tries to justify it by appeal to an argument form accepted premises. Considerations of closure of knowledge under entailment are typically introduced in epistemology only when relevant information is implied by one's beliefs and one is simply not aware of this fact. When viewed as proper arguments, the situation of CS can hardly be considered as an instance of this type of situation.

¹⁷E.g. Fodor and Pylyshyn's (1988) seminal assessment of classical cognitive architecture vs. connectionism, where they point out the possibility "that nonrepresentational neurological states are interconnected in the ways described by Connectionist models *but that the representational states themselves are not*", since "it is perfectly possible to implement a *Classical* cognitive architecture in such a network?" (*ibid.*, p. 10).

There is a further option for reconciling CSs as arguments with the considerations on closure despite these objections, namely, embracing that it is really the *computer* that is arguing. But should we allow that the computer is literally arguing, even if it lacks the *intention* to prove anything? Could we seriously say that the computer ‘is already convinced’ that a given output will occur, and is trying to convince us by executing the steps in the simulation model in the same way that we sometimes want to convince others by providing arguments?

The bottom line is this: The evidenced possibility of reconstruction in terms of arguments speaks in favor of CSs respecting the *logic* of arguments. This may be understood perfectly well in terms of their *algorithmic* nature and the implied rigidity of the input-output relation, similar to truth preservation in deductively valid arguments. But the considerations above demonstrate that CSs forfeit the *pragmatics* of arguments. In this sense the computer ‘executing an argument’ is more of a metaphor, and Beisbart’s (2012) “slogan” that “[c]omputer simulations *are* arguments” (emph. added) is misleading.

This conclusion, however, only addresses the question of whether CSs *are*, strictly speaking, arguments. It does not directly address the validity of EPT_A . To make contact with EPT_A we need to look at the *epistemology* of arguments.

Let us first ask how ‘epistemic power’ could be meant here, precisely. I see two ways: it could refer (i) to the *certainty* conveyed by the simulation, or (ii) to the *knowledge gain* that one receives from it. I will refer to these readings of EPT_A as $EPT_A(i)$ and $EPT_A(ii)$ below.

An objection to $EPT_A(i)$ arises in the context of the meta-arguments securing proximity between numerical and conceptual model in the *MC* case. Here the convergence is merely *probabilistic* (ultimately given by a law of large numbers; cf. James 1980, p. 1150):

In the statistical context, the ‘guarantee’ must be replaced by a statement of probability, so that the corresponding definition becomes: $A(n)$ is said to converge to B as n goes to infinity if for any probability $P[0 < P < 1]$, and any positive quantity δ , a k can be found such that for all $n > k$ the probability that $A(n)$ will be within δ of B is greater than P . Note that this is quite weak, in that no matter how big n is, $A(n)$ can never be guaranteed to be within a given distance of B . (James 1980, p. 1151)

But P , providing the probable proximity, expresses an *uninterpreted* probability, and one may wonder what P *should* be taken to signify. In fact, Beisbart and Norton (2012, p. 414) avoid the “thorny problem of explicating probability”, but claim that “an objective notion is used” which “assures us of a close association between the relative frequencies of random numbers and the corresponding probabilities for large sets of random numbers.”

A brief glance at the literature suffices to see that this is not an uncontroversial move: it is unclear whether “the second order probabilities” expressed by P are “to be understood subjectively or objectively” (Ben-Menahem and Hemmo 2012, p. 4), as “without any independent meaning given to P ,” claims from a law of large numbers to the connection between finite relative frequency and objective or subjective probability are “empty.” (Howson and Urbach 2006, p. 48) Put frankly, “the law of

large numbers is a mathematical truth which makes sense in any interpretation of probabilities.” (Schurz 2014, p. 153)

Assume now that P is taken to quantify the degrees of belief of an ideal, logically omniscient, and perfectly rational agent. Then such probabilistic convergence theorems can be fleshed out to say that this agent would assign degree of belief P to the frequencies in question approximating p to the specified amount. This provides a mark of rationality for non-ideal agents, given certain background assumptions about how to express and quantify rationality and belief, but nothing beyond that.

A second objection to $EPT_A(i)$ arises from the involvement of *inductive* inferences, to which Beisbart and Norton (cf. 2012, pp. 409, 411) refer as “preserving of the probability of truth”. This, first off, strikes me as a confusion between two uses of ‘induction’: inferences in which probability is preserved are strictly speaking the subject matter of *probability logic*, which is also sometimes called ‘inductive’. This goes against the “dominant position”, though, “that probability logic entirely belongs to deductive logic, and hence should not be concerned with inductive reasoning.” (Demey et al. 2013, sect. 1) This dominant position is quite sensible: the concluded probability assertions themselves (which in turn concern the probability of truth of other statements) are inferred with deductive validity from other probability assertions. In this sense – and in this sense only – ‘inductive’ inferences can preserve the probability of truth.

Proper inductive inferences are all strictly weaker.¹⁸ The assertion that a certain technique has worked m out of n times ($m < n$) may be true with probability 1; the assertion that it will do so the next time can only be concluded to be true with probability no greater than $m/n < 1$ by proper induction.

This latter sort of reasoning is, moreover, certainly entertained to bridge the gap between numerical and simulation model in many (if not most) cases. Here is Winsberg (2010, p. 122):

the techniques that simulationists use¹⁹ are ‘self-vindicating’[...]; whenever they make predictions or produce engineering accomplishments—their credibility as reliable techniques or reasonable assumptions grows. [T]hey carry with them their own history of prior success and accomplishments, and, when properly used, they can bring to the table independent warrant for belief in the models they are used to build.

Past experience with these simulation methods in various tests and applications will hence usually justify the replacement of a numerical technique by a particular non-isomorphic simulation technique. This is just a(n) (statistical) inductive prediction: given the past success of some simulation method (in m out of n cases), why would we doubt its future success (to a degree greater than $1 - m/n$)? But this has nothing to do with probabilistic convergence theorems or probability preservation.

¹⁸Cf. also Schurz (2014, p. 51) for a review of different kinds of induction.

¹⁹Cf. also Winsberg (2010, pp. 46 ff.) for an example of a specific simulation technique that, while connected to a discretization of the Euler equations, in no sense ‘directly derives’ from this discretization.

The point is this: The involvement of probabilistic convergence and ‘induction proper’ makes the knowledge gained from outcomes of simulations *less certain* than any conclusion of a deductively valid argument, e.g. from certain probabilistic-dynamical assumptions to the expected average behavior of a system. Since *that* conclusion is usually the statement of interest, I think that $EPT_A(i)$ is false.

An objection to $EPT_A(ii)$ transpires from the possibility of ‘psychological novelty’ or ‘surprise’ in CSs, as acknowledged by Beisbart (2012) and Morgan (2003) respectively. How, if one is in charge of the model implemented, and if the connection between input and output is rigid in a CS, can such unexpected new knowledge be understood? Kripke (1972, p. 159; emphasis added) famously noted that “one can learn a mathematical truth *a posteriori* by consulting a computing machine, or even by asking a mathematician.” This seems to apply perfectly in the case of CSs: one has to appeal to experience (look at the screen) to see the result of the simulation, even though that result is already logically contained in the simulation model. The knowledge gained from the output of a CS should hence be seen as *analytic a posteriori knowledge*;²⁰ the knowledge gained from the execution of an argument is *analytic a priori*.

What makes this interesting for the epistemology of simulation is that, for many of the complex simulations involved in HEP and other fields, one may realistically estimate that the entirety of humankind could probably not execute all the intricacies of certain simulations within the remaining lifetime of the earth, maybe even the universe. While every single step could be executed by hand, and sometimes even theorems exist which ensure the existence of solutions, it seems safe to say that we are often bound to gaining the desired knowledge in an *a posteriori* fashion, even though this knowledge is basically ‘analytic’, on account of one’s accepted models.²¹ Hence I think that $EPT_A(ii)$ is false as well: The knowledge gain from a CS may be vastly greater than that from any inference that could ever be drawn by human beings.

²⁰Note that this strictly concerns only the relation between simulation model and output: an analytic connection between conceptual or numerical model and output is typically undermined by the considerations on induction and convergence.

²¹Parker (Evidence and knowledge from computer simulation, unpublished manuscript) provides a detailed account of how CSs can produce ‘genuinely’ new knowledge, even if very high standards are imposed on such genuine novelty, including that a proposition can only count as genuinely new knowledge if it is not entailed by one’s beliefs. She concludes that for the typical scientist (‘Sam’), who may not be aware of (all of) the details of a given simulation model, the result of a simulation, when interpreted as an approximation to another proposition p , can lead to the genuinely new knowledge that a value specified in p lies within some margin of error. p itself might be entailed by one’s existing belief set, but Sam (and other scientists) might be unaware of the entailment, and of p . I think Parker’s analysis is compatible with my views and implies a nice addition: in HEP, any working scientist is like Sam, because of the complexity of the code and the fact that it is written and edited by many authors in concert. So the result of a simulation model, while ‘analytic’ on account of the entire code, is entailed by no single scientist’s belief set. Given the connections between the different models proposed in Section 1.2, and the relation between results from these, high energy physicists can thus in principle gain genuinely new knowledge from their CSs in the way outlined by Parker.

4 A positive account

4.1 CSs as surrogate experiments

Emphatically, I have only argued that one should reject EPT_E , EPT_A , and the notion that CSs *are* arguments—but *not* that one should reject the notion that they *are* experiments. This is so because I do not believe this rejection to be just.

Now Beisbart and Norton's claim "That two things can take the same role does not show them to be identical" (their p. 418) I find entirely correct – one should give independent reasons as to why one thinks of CSs as experiments. I agree, in other words, that it is wrong to think that "Monte Carlo simulations must be experiments *because* they can take the same *role* as experiments." (Beisbart and Norton 2012, p. 418; *emph. added*) But this merely says that CSs replacing experiments on target systems is the context of discovery of the hypothesis that they are experiments, not the context of its justification.

What could such justification look like? First off, note that Beisbart and Norton (2012, p. 411) argue for their argument view by claiming "that Monte Carlo simulations are powered epistemically either by [...] 'discovery' or by [...] 'inference' [...]", and that the former case is excluded. Inference, according to them, "requires no contact with the world and [...] transforms knowledge of the world already gained", whereas discovery "goes directly to the world" and is powered by "novel experience[...]" (Beisbart and Norton 2012, pp. 408-9) So, the argument goes, CSs are not powered by discovery because "[a] method only includes discovery if hitherto unknown properties characteristic of a particular physical system are recorded", and even in a MCS, "[e]verything that matters epistemically is already known in advance, namely that the random numbers follow a certain distribution."

Now assuming that discovery goes *directly* to the world rules out virtually all modern experimental discoveries as instances of 'discovery': Recall Morrison's careful discussion of measurements and the involvement of models, functioning as "mediating instruments" (Morrison and Morgan 1999) therein. Moreover, that "[e]verything that matters epistemically" are "the random numbers" which "follow a certain distribution" seems equally false to me: What matters most, epistemically, is the *output*, and since this output is often inevitably found out a posteriori, it is not true that "[e]verything that matters epistemically is already known in advance".

Thus pace Beisbart and Norton (2012), I here want to defend two claims: (a) CSs are powered by both inference *and* discovery, as are traditional laboratory experiments (cf. Morrison 2009); and (b) it is profitable to view CSs as experiments, even if EPT_E is ultimately incorrect. (a) should be understood as involving two sub-claims: (a.i) that CSs are powered by discovery at all, and (a.ii) that this discovery concerns the target system, not merely the models used or the simulation system. So should (b): (b.i) that CSs can be understood as experiments, and (b.ii) that it is profitable to do so.

(a.i) was actually already defended above: I have rejected the notion that discovery needs to go directly to the world, and argued that the access we have to the

output is often times inevitably a posteriori. So CSs *are* powered by novel experience (discovery). To argue for (a.ii), I will first have to spend some thoughts on (b.i) though.

Recall how experiments are typically characterized as observational situations that involve greater *control* than do field observations (cf. Schurz 2014, p. 35; Radder 2009, p. 3), and where this control can (partly) be executed in the form of *interventions* (cf. Radder 2009, p. 2). These interventions may influence the outcome, but must still allow genuine (albeit not necessarily unperturbed) observations in response. Certainly, an experiment is an *observational* situation, so it is at least partly also defined by the fact that we here attempt to gain knowledge by (indispensable) appeal to experience, not by pure thought alone.²²

Now think of a digital computer configured according to some simulation model. Then we could exert control over the computer via that model, and intervene on it by changing the values of the model's parameters (strictly speaking: by pushing the buttons on the keyboard and thereby applying electrical currents). In consequence, we would observe changes in the output (strictly speaking: in the behavior of the pixels on the computer screen), which would in the first place convey insights about the computer under these specific circumstances. Described in this fashion, we can understand a CS as an experiment *on the computer*, where the simulation model defines the parameters of the experiment. This provides sufficient grounds for accepting (b.i).

Parker (2009, p. 488) has given a very similar account of CSs as experiments. She identifies a CS as “an experiment in which the system intervened on is a programmed digital computer.” (ibid.) But usually the computer itself is *not* our target system; we use it to get insights into elementary particles, the weather, distant galaxies, the economy on a large scale, and so forth. Let me hence propose what I call the *surrogacy thesis*:

(ST) Computer simulations are *surrogate experiments*, executed consciously on the wrong kind of system, because findings on that (surrogate) system can be *mapped* to the target system, and because the surrogate can be *handled* in a way that the target cannot.

Clearly the fact that simulation model and output can somehow be mapped to a model of the target system and observations made on it is a necessary condition for CSs figuring as surrogate experiments. When talking of ‘mappings’, I here have in mind homomorphisms between the best models of target and simulating system, loosely following Winsberg's (2009) and Dardashti et al.'s (2015) appraisal of Parker's proposal.²³ ‘Best’ here really means two different things for target and simulating system though: The best model of the target will be one that fares best

²²Cf. also Beisbart (2018, p. 176) for a similar assessment.

²³In contrast to Winsberg and Dardashti et al. I talk of homomorphism rather than isomorphisms because it may not always be realistic that the mapping between the model of the target and the model of the computer is one-to-one; a single feature of the target may be represented (jointly) by multiple features of the computer. This all, of course, depends on the grain and the details of both models.

regarding all empirical evidence about it, and maybe also some practical considerations such as tractability or economy. The best model of the simulating system will, however, in contrast allow us to *view* the simulating system *as* a surrogate; it may disregard many know details, and focus largely on the formal properties of the simulating system (the computer), meaning properties “that can only be instantiated by a relation between/among other properties” (Parker 2009, p. 487), i.e. higher order relational properties.

If we establish such a homomorphic mapping (not necessarily explicitly), this will allow us to draw inferences (not necessarily deductively valid ones) from findings on the simulating system to properties of the target.²⁴ To demonstrate that we *can* so draw inferences to our target systems, it suffices to show that this has been done successfully in the past. Hence consider, as an example, Hughes’ (1999) detailed analysis of MCSs based on the Ising model and the inferences drawn from these. Such CSs, when understood as modeling atomic spins, lead to the insight that “critical behaviour at large length scales may be independent of seemingly important features of the atomic processes that give rise to it.” (ibid., p. 141) This clearly concerns systems other than digital computers. To count as a genuine insight about ferromagnetic systems, the results of course had to be consolidated with other experimental and theoretical results (cf. Hughes 1999, pp. 113-4). But that is inessential: any experimental result will have to be cross-checked by other experiments, typically involving different equipment and different techniques for analysis. The Ising model *did* lead to genuine insights about the relevant class of physical systems, even though this could not have been known without consolidation by other sources. Since this is an excellent example of how CSs were used to gain genuine insights about real world systems other than digital computers, it reasonably justifies acceptance of (a.ii).

Now to finally justify (b.ii), let me summarize, in brief, a couple of merits that I see associated with ST:

- (I) As Hughes’ example shows, ST is useful for understanding *how* CSs can lead to genuine, new knowledge: if we can sensibly map simulation model and output back to the target system, and if the insights thus gained can be consolidated by independent sources, we are somewhat justified in “relying on the physical processes at work within it [the computer – FJB].” (Hughes 1999, p. 139)
- (II) ST allows to view CSs in *continuity* to other types of simulation, such as analog simulation or table top experiments. Dardashti et al. (2015, p. 13), for instance, provide a detailed analysis of how to draw inferences in analog simulations,²⁵ and then (p. 14) find that

²⁴I will here not go into details about the precise form of these inferences, but elsewhere (ref. blinded) I provide a detailed analysis for a well-defined range of scientific contexts, as do Dardashti et al. (2015) for a different one.

²⁵By this term Dardashti et al. (2015, cf. p. 13) mean the pertaining of a syntactic isomorphism between the best models of two physical systems that can be exploited to obtain information about one system by experimenting on the other. Cf. also Winsberg (2010) on this issue.

the main difference between computer simulation and analogue simulation is simply that in computer simulation, the [simulating system – FJB] is a programmable digital computer, and the reasons that it meets the conditions articulated in [one step of the inference – FJB] is that it has been programmed precisely so as to meet those conditions.

- (III) The involvement of homomorphic mappings between models of both systems, required on account of ST, helps us to understand how CSs can be used to *explain* features of target systems: reasoning from *analogies* has been identified as one important kind of explanatory reasoning (e.g. Hesse 1966; Bartha 2010, p. 24), and analogies are typically spelled out by appeal to homomorphisms (e.g. Bartha 2010; Gentner 1983; cf. also Boge (How to infer explanations from simulated experiments, unpublished manuscript) for a detailed analysis of analogical reasoning from CSs in explanatory contexts).
- (IV) While ST does not fall prey to the arguments raised against EPT_E , as it involves considerably weaker claims, it still allows for an appreciation of typical *uses* of CSs: Hillerbrand (2013, pp. 62-3) e.g. identifies *complexity* and experimental *inaccessibility* as the main reasons to use CSs in empirical science. Both of these reasons boil down to the fact that the computer can be *handled* in a way that the target cannot.

All this taken together reasonably justifies (b.ii).

4.2 Surrogate experiments or surrogates for experiments?

Before concluding, let me contrast ST with, and defend it against, a quite recent proposal by Beisbart (2018) that bears some similarity but is ultimately at odds. Beisbart (2018, p. 192 ff.) namely urges to think of CSs as “modeled experiments”, meaning that simulation models “define a fictional system, a substitute the behavior of which is used to represent the behavior of the target” (p. 193) and on which “quasi-intervention” and “quasi-observation” are possible, the former meaning that “the simulationist can set the initial conditions and the values of important parameters[...] similar to manipulation and activities of control on the part of the experimenter in an experiment”, the latter that “the simulationist can take notice of the outputs from her simulation[.]” (p. 194)

There is an obvious sense in which Beisbart’s proposal is related to ST: the ‘fictional’ system defined by the simulation model is used as a “surrogate” (Beisbart 2018, e.g. pp. 193 and 195) for the target system; the surrogate is (quasi-)intervened on by manipulation of the parameters of a simulation model; and this somehow accounts for the generation of possible new knowledge about the target (cf. p. 199).

However, there are multiple respects in which both proposals are straightforwardly incompatible: the entire activity of ‘quasi-intervening’ and ‘quasi-observing’ is understood as a surrogate *for an experiment* (cf. p. 199), not as an experiment on a surrogate system; and Beisbart explicitly rejects the notion that we ‘observe the hardware’ in using CSs, whereas I endorse that we do make actual observations on the physical system that is the computer.

To fully appreciate the difference, let me spell out in more detail what I mean by the claim that we make observations on the computer. Note first that, like Beisbart (2018, cf. p. 176) and the rich literature he builds on, I endorse that intervention on, and observation of a system X are necessary conditions for an activity to count as an experiment on X . And like Parker (2009) I believe that this fact can be exploited to understand CSs as a species of experiment. To understand the suggestion in detail, note, moreover, that any computer, digital or analog, is just a complex physical system, more specifically, a complex, nested electrical circuit. Pushing the buttons on a (certain kind of) computer keyboard, for instance, means closing a tiny electrical circuit, embedded in a grid of wires, thereby effecting an electrical current. The location of the circuit-closure on the grid of wires can then be understood as a bit of information that encodes which key on the keyboard was pressed (cf. Clements 2006, pp. 437–40).

The central processing unit of the computer essentially operates on the basis of *gates* and *flip-flops* (cf. *ibid.*, pp. 25 ff. and 294 ff.). Gates are tiny circuits that will only accept or transmit signals in (i.e., react sensitively to) two specified voltage ranges. One of these can then be interpreted as a ‘1’, the other as a ‘0’, and the computer can thus physically realize logical operations such as ‘and’ or ‘not’. Flip-flops, on the other hand, are “sequential circuits” whose “output can remain in one of two stable states indefinitely, even if the input changes.” (*ibid.*, p. 102) They hence function as (information) storage devices.

The computer’s main storage devices are hard disk- or solid state drive and the random access memory. They too are realized in terms of electrical circuits. Dynamic random access memory (or DRAM), for instance, the most widespread type in modern desktop computers, stores data in the form of electric charges in the inter-electrode capacitance of a field effect transistor (cf. Clements 2006, p. 497), an arrangement of two negatively and one positively doped semi-conductors in which a significant current can only flow if a sufficient voltage is applied (e.g. Gross and Marx 2012, p. 550, for details). The DRAM is dynamic in that the “charge gradually leaks away” and will have to be restored “every 2 to 16 ms in an operation known as memory refreshing.” (Clements 2006, p. 497; *emph. omit.*) This has drawbacks regarding the interface with the CPU, but is cost wise preferable to other types of RAM since only one transistor is required.

Finally, there is the computer display, which today usually operates by appeal to liquid crystals, i.e., liquids whose molecules arrange in a fashion that is in some respects similar to certain solid state crystals (cf. *ibid.*, p. 448). In particular, arrangements of specific liquid crystals, sandwiched between two orthogonal polarization filters, can be used to precisely modify the amount of light transmitted by applying an electric field that changes the molecules’ orientations and allows them to alter the incident light’s polarization. Such an arrangement constitutes a cell or ‘pixel’ of a liquid crystal display (cf. *ibid.*, pp. 447–50; and cf. p. 459 therein for details on the implementation of different colors).

Programming a digital computer in a specific way means exerting control over these physical conditions. This part of the execution of a CS should be compared to the *preparation* stage of an experiment, as identified by Lange (2003, pp. 121–2), in

which we bring a certain object under study (in this case the main hardware of the computer, i.e., motherboard, CPU, hard drive, and so forth) into a specific relation to relevant equipment (in this case: keyboard, display, and pointing devices). The relation between the object and the equipment can then be exploited during a *run* of the experiment, a sequence of events leading from *initiation* to obtainment of what I want to call an *individual* outcome. Initiation and obtainment steps may involve taking action on the object via the equipment, i.e., the aforementioned interventions (cf. also Lange 2003, p. 122). In our case the initiation means a highly specific combination of button pushes on, and, in the latter case, movements of, keyboard and pointing device (e.g. mouse). This in turn means nothing but the application of a specific series of electrical impulses in the form of currents and voltage changes to the main hardware of the computer. The obtainment of an individual outcome hence means observing, after some period of time, the effects of the resulting changes to the main hardware on the configuration of the display, i.e., the (sequence of) changes the cells of liquid crystals undergo as a consequence of their interaction with the main hardware.

Typically multiple runs of an experiment and additional statistical modeling steps will be necessary to obtain a suitable ‘result’ or *general* outcome, such as an expectation value or a set of probabilities for specified individual outcomes.²⁶ This too is relevant in the case of CSs, as should be obvious for MCSs²⁷ but can be justified for deterministic CSs as well: The the “rule of movement” (Schelling 1971, p. 154) in Schelling’s famous model of segregation, for instance, is perfectly deterministic; but establishing a serious result relating segregation to the preferences of individual agents from corresponding simulation studies requires several runs and the statement of an expected segregation pattern. The fact that the acquisition of the relevant individual outcomes is provided by (direct) *observation*, however, remains unimpaired by this.

Now observation is itself of course a tricky concept. Shapere (1982, cf. p. 492), for instance, famously suggested to extend the narrower philosophical concept that is closely tied to sense perception to a (more or less) unperturbed transmission of information from a source to some appropriate receptor. His reason was that one would otherwise have a hard time making sense of the term’s usage in science; specifically astrophysicists’ claims to ‘direct observation’ of the sun’s interior by means of neutrinos (cf. *ibid.*, pp. 485-91). My intention here is neither to provide a detailed account of observation nor to evaluate Shapere’s account in detail.²⁸ I rather want to suggest that, irrespective of the specific details of one’s account of observation, looking at a computer display and seeing pixels flash should count as an ‘observation’. What the considerations of Shapere and others (e.g. Hacking 1983, Chaps. 10-11) allow me to say in addition is what these observations are observations *of*: they should count

²⁶E.g. Mayo (1996, p. 157) or Karaca (2017, p. 12) in this context.

²⁷Compare, in this connection, the set of N reconstructing arguments, as specified in Section 2.3, required to obtain the reconstructing argument for the set of probability distributions.

²⁸E.g. Falkenburg (2007, pp. 66 ff.) for a critical appraisal and suggested improvement of Shapere’s account.

as observations of the changes effected in the computer's main hardware in consequence of the initial interventions undertaken with keyboard and pointing device; and the computer display figures as an instrument in these observations.²⁹

In short, my proposal here is this: we prepare the computer by programming it; we initiate a run of an experiment by intervening on it with mouse and keyboard; and we then observe the changes it undergoes by using its display as an instrument of observation. This procedure is typically executed multiple times or in a way that naturally segments into multiple runs, and a result (or general outcome) is then abstracted from all the individual outcomes of these runs.

The precise goings on of the hardware are, however, almost unequivocally not what one is interested in. The relation between this result and the behavior of the system one is actually interested in—and with it the purpose of the experiment on the computer just described—is mediated by the simulation model, which allows to establish a suitable morphism with the (best model of the) target system. This is exactly the content of ST.³⁰ It is the remarkable flexibility of digital computers to serve, via appropriate programming, as surrogates for a huge range of vastly different (target) systems that makes them such important tools for modern science.

My assessment should be contrasted with Beisbart's (2018) own more specific considerations on experiments and their relation to CSs. In particular, a central element in his approach to experimentation is the imposition of an additional necessary condition, that for some procedure executed on system *X* to count as an experiment on *X* one must execute interventions and observations on *X* "with the superordinate aim to obtain information about the way *X* behaves and reacts to the intervention." (Beisbart 2018, p. 176) In effect, these considerations serve to demarcate intervention and observation against the quasi-versions invoked by Beisbart: In contrast to the 'real' activities, the latter "are done with the superordinate aim of learning about the system on which an experiment is *modeled*." (Beisbart 2018, p. 195; *emph. added*) This he considers "parallel to", but at the same time clearly distinct from, "experiments, in which intervention and observation are done with the superordinate aim of learning about the *object of the experiment* [...]." (*ibid.*, *emph. added*)

These axiological considerations are obviously crucial for distinguishing the quasi- from the 'real' activities, and thus for identifying CSs merely as quasi-experiments, well-distinguished from actual experiments. Beisbart's additional necessary condition, however, and with it the force of the associated argument, is rather doubt worthy. For consider, say, pharmaceutical experiments on *laboratory animals*, standardly (and in my opinion sadly) considered suitable 'model organisms' for

²⁹What is meant by 'instrument' here? That is a similarly intricate question as that to the concept of observation, whence I will refrain from a detailed analysis in this case as well. As a minimal working definition, however, I suggest that in the context of experimentation an instrument is a piece of equipment that *correlates* with the object under study in way that is exploitable for the task at hand (e.g.: observation). This view is compatible with the more detailed accounts given e.g. by Harré (2003, cf. p. 20) or Heidelberger (2003, cf. p. 147).

³⁰It is clear that this is not the description a scientist using a CS would give when asked what she was doing. I will address this issue a little below, as it actually constitutes another objection raised by Beisbart.

human beings. Administering some drug to a group of mice with highly specific features due to breeding under highly controlled laboratory conditions and observing the consecutive changes to their health is undoubtedly a kind of experiment, and undoubtedly an experiment *on the mice*. But the superordinate aim of such an experiment is certainly not to learn something about the mice themselves. Rather, the superordinate aim is to learn something about the potential effects of the drug on *human* health—i.e., about a *different* (class of) system(s).

Beisbart's necessary condition is hence implausible: it can be used to rule out clear cases of experimentation as instances of the very same. A fortiori, it also does not deliver a good reason to *not* think of CSs as experiments on surrogate systems, involving intervention and observation proper; for accepting it would immediately rule out experiments on laboratory mice as proper experiments as well—an unacceptable consequence. Learning about the system experimented on may, in other words, sometimes be a merely *subordinate* aim.

In contrast, ST here suggests that there are in fact commonalities between both cases: experimenting on mice to obtain potential information about a substance's effect on human health means using a surrogate system ('model organism') to perform an experiment/obtain information that would otherwise be (legally) forestalled/unavailable. And in both cases the quality of the information transfer depends on one's ability to reliably map the results back to the target system; a fact that has been known to cause problems in the case of animal studies (e.g. Pound et al. 2004). ST thus exhibits a *unifying power* that is absent from Beisbart's proposal, as it allows to view different scientific practices as instances of a common methodology; a fact was already indicated in merit (II), discussed in the last section, and that I shall make more prominent below.

There is another motivating reason for rejecting a view such as ST, which is given by Beisbart in terms of the following thought experiment:

Suppose that [we give] a mathematical genius who can do every calculation we wish her to do in a few milliseconds[...] an algorithm that evaluates how certain physical characteristics behave as functions of time according to a model. [...] We could use these results in exactly the same way in which we could use results from a CS that follows the same algorithm. [...] But [...] this does not constitute an experiment. It is obviously not an experiment on the hardware of a computer (or an experiment on the brain). (Beisbart 2018, p. 192)

The thought experiment is clearly intended as a *reductio*, and it has some *prima facie* appeal. But there are still many asymmetries between the situation of the genius and ones in which we use a CS. Instructing the genius with a set of algorithms, she is, for instance, at liberty to use a different set, e.g. because, being a genius, she understands the problem extraordinarily well and finds those algorithms more convenient. Thus with the genius, the algorithms do not convey the sort of control that they do with CSs, which was one crucial reason for thinking of the latter as experiments.

Moreover, should we find that the genius makes a false claim as a result of her calculations, we would certainly not suspect that there is something fundamentally wrong with her brain. In contrast, when we find an unexpected result in a CS, we

may well wonder whether this is due to fundamental limitations implied by the hardware. Brain and computer hardware play markedly different roles in both kinds of scenarios. The case that we do in fact observe / experiment on the behavior of the computer hardware, in the specific circumstances defined by the implementation of a given simulation model on a given digital computer, can hence be made in spite of Beisbart's attempted reductio.

Finally, a central concern of Beisbart's (2018) is that viewing CSs as experiments is "unnatural". It is debatable (though not unequivocally clear) that this holds true from the individual researcher's perspective, who might not think of herself as performing an experiment on a computer when executing a CS. But there is nothing wrong with taking a bird's eye perspective here and comparing the similarities between traditional experiments and situations in which CSs are used. One may then find, as I have done above, that from this perspective the view of CSs as a species of experiment is perfectly 'natural'.

To see the issue more clearly, consider merit (II) that I had argued to be associated with ST in Section 4.1, the continuity it implies between CSs and other kinds of simulation.³¹ If we view the practice of using CSs as a surrogate for an experiment, not as performing an experiment on a surrogate system, this would imply that CSs are not simulations in the same sense as analog simulations or, say, human-in-the-loop simulations.

For analog simulations this should be rather obvious: Just consider Dardashti et al.'s example of dumb holes simulating black holes, and the involvement of (iso)morphisms therein (cf. their pp. 11 ff.). It seems quite artificial to interpret the interventions on the vat of fluid in which dumb holes can occur as 'quasi-interventions', and to also model these in order to be able to map the situation back to a conceivable experiment on a black hole. And it is also far from clear how to even *model* possible interventions on a black hole.

Human-in-the-loop simulations, on the other hand, are simulations in which a human being is "reacting to inputs from other simulation components, and generating outputs that affect the course of simulation." (Folds 2015, p. 175) For instance, one might want to assess "the suitability of an in-vehicle collision warning system for an automobile," and here "the presence of a driver to receive and respond to the alerts is key to evaluating the overall system performance." (ibid., p. 176) If this was viewed as a surrogate for an experiment, not as an experiment on a surrogate system (driver + stand-in for automobile, driving environment, and warning system), it would mean that the experimenter was not directly interested in the consequences of drivers' using a particular warning system, but rather in how an experiment involving drivers in automobiles with warning systems would turn out. This too seems quite artificial.

If we accept Beisbart's view of CSs, we are hence forced to accept that in 'computer simulation' we mean something entirely different by 'simulation' than for the other two kinds. Now why should CSs be simulations in such a rather different sense?

³¹I admit that the other merits discussed therein might be adapted to Beisbart's account, albeit with some ifs and buts.

Where, more precisely, is the *cut-off* between the two? Here is a famous quote by Feynman et al. (2013, p. 25-14–25-15; *emph. omit.*):

Suppose we have designed an automobile, and we want to know how much it is going to shake when it goes over a certain kind of bumpy road. We build an electrical circuit with inductances to represent the inertia of the wheels, spring constants as capacitances to represent the springs of the wheels, resistors to represent the shock absorbers, and so on for the other parts of the automobile. [...] This is called an analog computer. It [...] imitates the problem that we want to solve by making another problem, which has the same equation, but in another circumstance of nature[.]

The situation can be understood quite directly as a case of analog simulation, because the analog computer structurally obeys the same laws. At first glance it thus constitutes special case of simulation using computers because analog computers operate by appeal to continuous variables whereas most modern computers are digital and store and transmit information in discrete units (cf. above). But recall that a digital computer too is ultimately just a highly complex electrical circuit, and recall also that Dardashti et al.'s analysis presupposes appropriate *modeling* frameworks. Given what was said about the 'best' models in this context in Section 4.1, the following question cries out for an answer: At what level of complexity can a circuit not be modeled in such a way as to function as a simulating system used in a surrogate experiment anymore, but must be seen as the carrier of a fictional system in a surrogate *for* an experiment? There does not seem to be a non-arbitrary answer.

As I have stressed above, we can describe the activity of simulating with the aid of a programmed digital computer perfectly well in the terms standardly describing conventional experiments. It seems inessential, moreover, that an individual researcher might not think of herself as performing an experiment in the sense established. For compare her situation to that of a member of some particular religious group. A (strong) case can be made (cf. Wilson 2005), that religious beliefs have been selected throughout social and biological evolution because of their functions in regulating social life, particularly the sustaining and sharing of food sources by members of the respective religious groups. In executing religious practices, however, the religious practitioner will certainly not think of herself as merely sustaining food sources for herself and others. Rather, she will see herself as serving an omnipotent god or other transcendent personal entity. This does not change the fact that, from the point of view of the evolutionist scientist, effectively sustaining and sharing food sources is most likely all she is really doing.

Clearly the case is not perfectly analogous to that of the scientist using a CS, who will probably not have a firm belief in the existence of the fictional or virtual entities that could be said to exist on account of the simulation model. But there is still enough of an analogy to establish what I have in mind: that it is irrelevant to the appropriate or most natural description of an activity, when studied from the outside, whether the participants in that activity think of themselves in terms of that description or not. I conclude that while Beisbart's account may be 'natural from within', i.e. from the point of view of the individual simulationist, this is not so from the philosopher's

bird's eye perspective; among other things because it implies an arbitrary distinction between CSs and other kinds of simulation – a feature that is absent from ST.

5 Summary and outlook

In this paper I have argued that CSs are not arguments, but rather surrogate experiments. To do so, I have provided a detailed analysis of two antipodal views of CSs, namely, of CSs as arguments and of CSs as experiments with an epistemic stature sometimes equal to that of a corresponding laboratory experiment. I have then distinguished the reconstruction and replacement theses from the respective epistemic power theses, EPT_A and EPT_E , and demonstrated that replacement and reconstruction can live in perfect harmony, whence evidence for them does not support the two (mutually exclusive) epistemological hypotheses.

I have then argued that (a.i) CSs are not just powered by inference, but also by discovery, and (a.ii) that this discovery concerns a target system, not just the computer. I have also argued that (b.i) CSs can be viewed as experiments, and (b.ii) profitably so. All this together supports my surrogacy thesis (ST).

Given that I have rejected EPT_A and EPT_E , I now owe an epistemic power thesis of my own, compatible with ST, and summarizing the insights gained from the discussion. For obvious reasons I shall call it EPT_S :

EPT_S The epistemic power of a CS is less than that of a corresponding laboratory experiment involving the target system, because of the possibility of confoundment in the latter. It is greater in terms of knowledge gain than that of a reconstructing argument, because of the possibility of surprise, but less in terms of certainty, because of the involvement of induction and, in the case of MCS, higher order probabilities.

Since the target system may be impossible to handle in such a way as to gain experimental insight into it, a suitable CS may increase one's ability to discover something about it. Since such discovery requires the drawing of inferences – often times not deductively valid ones – from a simulation output to the behavior of the target system, the epistemic power of the CS is that of these inferences.

This is a view of CSs consistent with Hughes (1999), Parker (2009), Winsberg (2009), and Dardashti et al. (2015). It should be clear that this view, while also a view of CSs as a kind of experiment, is at odds with Morrison's thinking to the extent that EPT_E is replaced by EPT_S , in favor of the arguments given for an in-principle epistemic priority of traditional experimentation over CSs. While also embracing a notion of 'surrogacy', moreover, it is equally at odds with Beisbart's (2018) view of CSs as surrogates for experiments; and I believe to have shown it to be safe from Beisbart's arguments and more natural from the philosopher's point of view. A task that remains is to classify the inferences in question precisely, an endeavor beyond the scope of this paper. Cf. however Dardashti et al. (2015) and Boge (How to infer explanations from simulated experiments, unpublished manuscript) for steps in this direction.

Acknowledgements The research for this paper was conducted as part of the research unit *The Epistemology of the Large Hadron Collider*, funded by the German Research Foundation (DFG; grant FOR 2063). I thank Dominnc Hirschbühl for help with the physics parts, Rafaela Hillerbrand, Gregor Schiemann, Christian Zeitnitz, Michael Krämer, and Paul Grünke for helpful comments on an earlier version of this paper, and Claus Beisbart for exchange on some details of his joint paper with John D. Norton and discussions at MS8. I would also like to thank two anonymous reviewers for detailed and helpful suggestions on improvements of the original manuscript, and one of them in particular for pushing me on certain issues the discussion of which has clearly improved the overall structure of the paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barberousse, A., Franceschelli, S., Imbert, C. (2009). Computer simulations as experiments. *Synthese*, 169(3), 557–574.
- Bartha, P. (2010). *By parallel reasoning Oxford*. New York: Oxford University Press.
- Beisbart, C. (2012). How can computer simulations produce new knowledge? *European Journal for Philosophy of Science*, 2(3), 395–434.
- Beisbart, C. (2018). Are computer simulations experiments? And if not, how are they related to each other? *European Journal for Philosophy of Science*, 8(2), 171–204.
- Beisbart, C., & Norton, J.D. (2012). Why Monte Carlo simulations are inferences and not experiments. *International Studies in the Philosophy of Science*, 26(4), 403–422.
- Ben-Menahem, Y., & Hemmo, M. (2012). *Probability in physics*. Berlin: Springer.
- Clements, A. (2006). *Principles of computer hardware*, 4th Edn. Oxford, New York: Oxford University Press.
- Cowan, G. (1998). *Statistical data analysis*. Oxford: Clarendon Press.
- Dardashti, R., Thébault, K.P., Winsberg, E. (2015). Confirmation via analogue simulation: what dumb holes could tell us about gravity. *The British Journal for the Philosophy of Science* (p. axv010).
- Demey, L., Kooi, B., Sack, J. (2013). Logic and probability. In E. N. Zalta (Ed.) *The Stanford encyclopedia of philosophy, the metaphysics research lab, center for the study of language and information (CSLI)*. Stanford University.
- Falkenburg, B. (2007). *Particle metaphysics. A critical account of subatomic reality*. Berlin: Springer.
- Feynman, R., Leighton, R., Sands, M. (2013). *The Feynman lectures on physics*, Desktop Edition Volume I. Basic Books.
- Fodor, J.A., & Pylyshyn, Z.W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition*, 28(1), 3–71.
- Folds, D.J. (2015). Human in the loop simulation. In M. L. Lorper (Ed.) *Modeling and simulation in the systems engineering life cycle* (pp. 175–183). London: Springer.
- Galison, P. (1996). Computer simulations and the trading zone. In P. Galison, & D. J. Stump (Eds.) *The disunity of science. Boundaries, contexts, and power* (pp. 118–157). Stanford University Press: Stanford.
- GEANT Collaboration (2016). Physics reference manual. *GEANT 4 Release*, 10, 4.
- Gentner, D. (1983). Structure-mapping: a theoretical framework for analogy. *Cognitive Science*, 7(2), 155–170.
- Giere, R.N. (2009). Is computer simulation changing the face of experimentation? *Philosophical Studies*, 143, 59–62.
- Gross, R., & Marx, A. (2012). *Festkörperphysik*. Munich: Oldenbourg Verlag.
- Hacking, I. (1983). *Representing and intervening Cambridge*. New York: Cambridge University Press.

- Harré, R. (2003). The materiality of Instruments in a metaphysics for experiments. In H. Radder (Ed.) *The philosophy of scientific experimentation* (pp. 19–38). Pittsburgh: The University of Pittsburgh Press.
- Heidelberger, M. (2003). Theory-ladenness and scientific instruments in experimentation. In H. Radder (Ed.) *The philosophy of scientific experimentation* (pp. 138–151). Pittsburgh: The University of Pittsburgh Press.
- Hesse, M.B. (1966). *Models and analogies in science*. Notre Dame: University of Notre Dame Press.
- Hillerbrand, R. (2013). Order out of chaos? A case study in high energy physics. *Studia Philosophica Estonica*, 5(2), 61–78.
- Howson, C., & Urbach, P. (2006). *Scientific reasoning the Bayesian approach*, 3rd Edn. Chicago, La Salle, Illinois: Open Court.
- Hughes, R.I.G. (1999). The ising model, computer simulation, and universal physics. In M. Morrison, & M. S. Morgan (Eds.) *Models as mediators* (pp. 97–145). Cambridge: Cambridge University Press.
- James, F. (1980). Monte Carlo theory and practice. *Reports on Progress in Physics*, 43(9), 1145.
- Karaca, K. (2017). Lessons from the large Hadron Collider for model-based experimentation: the concept of a model of data acquisition and the scope of the hierarchy of models. *Synthese*. <https://doi.org/10.1007/s11229-017-1453-5>.
- Krauth, W. (2006). *Statistical mechanics. Algorithms and computations*. Oxford: Oxford University Press.
- Kripke, S.A. (1972). *Naming and necessity*. Cambridge: Harvard University Press.
- Lange, R. (2003). Technology as basis and object of experimental practices. In H. Radder (Ed.) *The philosophy of scientific experimentation* (pp. 119–137). Pittsburgh: The University of Pittsburgh Press.
- Lista, L. (2017). *Statistical methods for data analysis in particle physics*. Cham: Springer International Publishing.
- Massimi, M., & Bhimji, W. (2015). Computer simulations and experiments: the case of the Higgs Boson. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 51, 71–81.
- Mayo, D. (1996). *Error and the growth of experimental knowledge*. Chicago: University of Chicago Press.
- Menary, R. (2010). Introduction: the extended mind in focus. In R. Menary (Ed.) *The extended mind*. Cambridge: MIT Press.
- Morgan, M. (2002). Model experiments and models in experiments. In L. Magnani, & N. Nersessian (Eds.) *Model-based reasoning: science, technology, values* (pp. 41–58). New York: Kluwer.
- Morgan, M. (2003). Experiments without material intervention: model experiments, virtual experiments and virtually experiments. In H. Radder (Ed.) *The philosophy of scientific experimentation* (pp. 216–235). Pittsburgh: University of Pittsburgh Press.
- Morgan, M.S. (2005). Experiments versus models: new phenomena, inference and surprise. *Journal of Economic Methodology*, 12(2), 317–329.
- Morrison, M. (2009). Models, measurement and computer simulation: the changing face of experimentation. *Philosophical Studies*, 143(1), 33–57.
- Morrison, M. (2015). *Reconstructing reality: models, mathematics, and simulations*. Oxford: Oxford University Press.
- Morrison, M., & Morgan, M.S. (1999). Models as mediating instruments. In M. Morrison, & M. S. Morgan (Eds.) *Models as mediators* (pp. 10–37). Cambridge: Cambridge University Press.
- Nagle, R., Saff, E., Snider, A. (2012). *Fundamentals of differential equations and boundary value problems*, 6th Edn. Pearson Education Ltd.
- Parker, W.S. (2009). Does matter really matter? Computer simulations, experiments, and materiality. *Synthese*, 169(3), 483–496.
- Parker, W.S. (2010). An instrument for what? Digital computers, simulation and scientific practice. *Spontaneous Generations*, 4(1), 39–44.
- Parker, W.S. (2017). Computer simulation, measurement, and data assimilation. *The British Journal for the Philosophy of Science*, 68(1), 273–304.
- Pound, P., Ebrahim, S., Sandercock, P., Bracken, M.B., Roberts, I. (2004). Where is the evidence that animal research benefits humans? *British Medical Journal*, 328(7438), 514–517.
- Radder, H. (2009). The philosophy of scientific experimentation: a review. *Automated Experimentation*, 1(1), 2.
- Saam, N.J. (2017). What is a computer simulation? A review of a passionate debate. *Journal for General Philosophy of Science*, 48, 293–309.

- Schelling, T.C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, 1(2), 143–186.
- Schurz, G. (2014). *Philosophy of science. A unified approach New York*. London: Routledge.
- Shapere, D. (1982). The concept of observation in science and philosophy. *Philosophy of Science*, 49(4), 485–525.
- Thijssen, J. (2007). *Computational physics*. Cambridge: Cambridge University Press.
- Wedgwood, R. (2006). The normative force of reasoning. *Noûs*, 40(4), 660–686.
- Wilson, D.S. (2005). Testing major evolutionary hypotheses about religion with a random sample. *Human Nature*, 16(4), 382–409. <https://doi.org/10.1007/s12110-005-1016-1>.
- Winsberg, E. (2009). A tale of two methods. *Synthese*, 169(3), 575–592.
- Winsberg, E. (2010). *Science in the age of computer simulation*. Chicago: University of Chicago Press.
- Winsberg, E. (2015). Computer simulations in science. In E. N. Zalta (Ed.) *The Stanford encyclopedia of philosophy, the metaphysics research lab, center for the study of language and information (CSLI)*. Stanford University.

Affiliations

Florian J. Boge^{1,2} 

- ¹ Interdisciplinary Centre for Science and Technology Studies (IZWT), Bergische Universität Wuppertal, Gaußstr. 20, 42119 Wuppertal, Germany
- ² Institute for Theoretical Particle Physics and Cosmology, RWTH Aachen University, Otto-Blumenthal-Straße, 52074 Aachen, Germany