

# **Inconsistent Idealizations and Inferentialism about Scientific Representation**

Peter Tan (Fordham University)

*Forthcoming in Studies in History and Philosophy of Science Part A.*

*This is a preprint. Please cite the official version.*

## ***Abstract***

Inferentialists about scientific representation hold that an apparatus's representing a target system consists in the apparatus allowing "surrogate inferences" about the target. I argue that a serious problem for inferentialism arises from the fact that many scientific theories and models contain internal inconsistencies. Inferentialism, left unamended, implies that inconsistent scientific models have unlimited representational power, since an inconsistency permits any conclusion to be inferred. I consider a number of ways that inferentialists can respond to this challenge before suggesting my own solution. I develop an analogy to exploitable glitches in a game. Even though inconsistent representational apparatuses may in some sense allow for contradictions to be generated within them, doing so violates the intended function of the apparatus's parts and hence violates representational "gameplay."

## **1 Introduction**

Scientific representations come in all sorts. One can use a concrete system of sticks and balls to represent a molecule, a pair of mathematical equations to represent an ecology, a squiggly line and some arrows drawn on a page to represent electromagnetic interactions between subatomic particles, and so on. The aim of an account of scientific representation to provide a *general* answer to the following question: what does it mean for any abstract or concrete apparatus to count as a scientific representation of a particular system?

Inferentialists about scientific representation hold that the representation relation between a representational apparatus like a model or simulation and its target system is constituted by the apparatus's ability to facilitate inferences of a certain kind about that target. By inferentialists' lights, the hallmark of scientific representation is not some particular mind-independent relation that an apparatus bears to its target, like isomorphism or similarity. Rather, what makes an apparatus a representation of a target system is the extent to which cognitive agents can use it to

make “surrogate inferences” about that target, so-called because those inferences require treating the apparatus as a surrogate for manipulating and intervening on the target.

This paper argues that inferentialism must be amended to avoid a serious problem. This problem arises from the observation that many representational apparatuses used in scientific practice are internally inconsistent, together with the simple fact that (classically) *any* inference can be drawn from an inconsistency. Inferentialists must thus amend their view so that only a particular *kind of* (or way of drawing) surrogate inference legitimately contributes to an apparatus’s representational capacity. I shall begin by introducing inferentialism in more detail. Then, I present the challenge for inferentialists, consider some possible responses, and argue for my own solution.

## 2 Inferentialism: A Primer

When we speak of scientific representation, which relation are we speaking of? Inferentialists answer: what is essential to and constitutive of scientific representation is that it is *inferential*. What scientific representation most fundamentally is, they say, is a capacity for surrogate inference and reasoning, viz., the kind of reasoning in which an apparatus is manipulated, intervened on, and so on, in place of its target.

To illustrate, consider a representational apparatus as a tiny world of its own, complete with its own laws and nomological structure, which can be manipulated and intervened on.<sup>1</sup> In the process of surrogate reasoning about some concrete target system using this tiny “world”, one first denotes elements of the target system with certain elements of the representational system, then determines what happens *in the representational system* when, say, a certain variable is manipulated. Finally, these events or facts within the representational system are interpreted in terms of the concrete system the apparatus denotes.<sup>2</sup> To call on a well-worn example, when one uses a subway map, one first denotes concrete objects with different parts of the map (colored lines

---

<sup>1</sup> This notion is originally due primarily to R.I.G Hughes’s (1997) “denotation-demonstration-interpretation” view of scientific representation, as well as Robert Sugden’s (2000, 2009) view of scientific models as “credible worlds”.

<sup>2</sup> Cf. Bueno and Colyvan (2011), de Donato Rodriguez and Zamora Bonilla (2009, p102-103), Hughes (1997), Sugden (2000, 2009), Suárez (2004, 2015), Swoyer (1991).

denote rail lines, white dots denote stops, etc.). Then, when you want to learn how to get from one stop to another, you do something *within the map*. A fact about what obtains *in the map* (the tracing of your finger, along a colored line, from one white dot to another) is then interpreted as a fact about the concrete target system whose parts have been denoted with the different elements in the map.

Inferentialism comes in many forms. Some inferentialists articulate versions of the view where an apparatus's capacity to be used in this surrogative capacity is that *in virtue of which* representations count as representations at all (Kuorikoski and Ylikoski 2015, p3827). Other inferentialists defend a more deflationary tone, on which there is nothing more substantive that one can say about the relation of scientific representation except that it occurs exactly when there is a capacity for surrogative reasoning (Suárez 2004, 2015). Others say that it is *interpretations* of those parts of a representation which allow for surrogative reasoning that truly render an apparatus representational (Contessa 2007). Whatever differences there may end up being, the key hallmark of all inferentialist views is that scientific representation is primarily characterized by its functional role (cf. Chakravartty 2010). That is, what makes a scientific representation is a capacity to facilitate certain kinds of cognitive activities or epistemic goals, and chiefly for inferentialists, the activity of surrogative inference.

When it comes to other accounts of scientific representation, inferentialists are happy to admit that features like qualitative or structural similarity can facilitate surrogative reasoning. As Suárez notes in his original defense of the inferential view, surrogative reasoning “requires [a model] to have the internal structure that allows informed agents to correctly draw [surrogative] inferences” (2004, p774). But what distinguishes inferentialism from other views that might more heavily emphasize these qualitative or structural relations is that these relations are viewed as mere means by which an apparatus allows for more or fewer inferences to be made. By inferentialists' lights, it is their common byproduct—the possibility of surrogative inferences regarding a target—that ultimately constitute the fact that the apparatus represents what it does.

### 3 The Challenge from Inconsistent Idealizations

I shall argue that inferentialism faces a serious challenge. The challenge shows that it cannot be the case that all that scientific representation amounts to, unqualifiedly, is inferential power. Instead, it must be a particular *kind* of surrogative inference (or way of drawing surrogative inferences) that constitutes representational capacity.

The challenge arises from the simple observation that representational apparatuses in science (theories, models, etc.) are sometimes logically inconsistent, or at least are idealized in logically inconsistent ways. It is well-known that competing models of the same phenomenon can often represent that phenomenon in logically incompatible ways, such as models of water that represent water either as continuous, or as composed of particles. But there are some *internally* inconsistent models and theories. A few plausible examples are:

- Bohr’s theory of the atom (a canonical example)
- The old quantum theory of blackbody radiation (see Norton 1987)
- Kirchoff’s diffraction theory (Saatsi and Vickers 2011, Buchwald and Yeang 2012)
- Lorentz’s early theory of the electron (Vickers 2013, §6)

In each of these cases, it is possible to derive both  $p$  and  $\neg p$  within the theory for some  $p$ . In at least a few of these examples, the theory or model is logically inconsistent because some assumptions within the theory require certain quantum mechanical principles to be true, while other assumptions require contradictory classical principles to be true. This seems to be the case with Bohr’s theory of the atom, which ends up representing orbiting electrons both as emitting radiation and *not* emitting radiation. It is also the case with the old quantum blackbody theory of radiation (Norton 1987). But whatever the basis of the inconsistency happens to be, these all seem to be plausible examples where a scientific theory or model taken as a whole contains a contradiction.<sup>3</sup>

---

<sup>3</sup> Note the phrase “taken as a whole”. Inferentialism claims, again, that a model’s representing a target consist in its facilitating surrogative inferences about that target; this is a claim about what entire models do or don’t facilitate. In a few moments (§4.2), I consider whether the inferentialist can respond to the specter of inconsistency by appealing to what specific *parts* of a model do or don’t facilitate. Additionally, I am aware that recent work on inconsistent scientific theories suggests that even if there is inconsistency in a theory “taken as a whole”, scientific theories or models are rarely inconsistent *in practice* or when understood correctly (see, e.g., Vickers 2013). Proposals on understanding *why* they are not inconsistent in practice inform my own solution, which relies on an analogy to legitimate moves in gameplay. I defend my solution in Section 5 of this paper.

Inferentialists say that the representational capacity of an apparatus *just is* its capacity to facilitate surrogative inferences, i.e., inferences from facts within a model to facts about its target. This cannot be true of all surrogative inferences unrestrictedly. For with an inconsistent representational apparatus, if all that it is for it to represent a target (and how well) is which surrogative inferences a cognitive agent can perform with it, a competent agent can draw literally *any* surrogative inference she likes. She can simply start with the fact that each of  $p$  and  $\neg p$  obtains somewhere within the representational system, infer their conjunction, and validly infer that some arbitrary  $q$  obtains. Voila! Since this can be done for any arbitrary  $q$ , an internally inconsistent apparatus has unlimited inferential power. And thus, by the inferentialist's lights, it also has unlimited representational power regarding its target, i.e., allows any inference whatsoever to be drawn about the target. Additionally, not only will an inconsistent apparatus allow its users to infer anything they like about their target for free, one can also use an inconsistent apparatus to represent any other target one likes. Simply denote the elements of any arbitrary concrete system with a few soon-to-be-inferred terms in the apparatus and infer away! If inferentialism is true, inconsistent representational apparatuses also *universally* represent.

Obviously, something is amiss. To be sure, inferring any arbitrary  $q$  from an apparatus's internal logical inconsistency is a classically valid inference. But it cannot plausibly be the kind of inference that constitutes the fact that an apparatus represents what it represents, with whatever fidelity of representation it does. There must be some restriction on which kinds of surrogative inferences actually count as constituting an apparatus's representational capacity. In other words, these kinds of "explosive" surrogative inferences (hereafter "explosive inferences"), that merely exploit a contradiction in an inconsistent model to infer something, must be ruled out. Inferentialism must be revised such that representation happens not simply only when apparatus facilitates surrogative inferences regarding a target, but such that only a particular kind of surrogative inference, or some particular way of performing surrogative inferences is representation-constituting.

## **4 Some Preliminary Solutions**

### **4.1 Consistent Sub-Representations?**

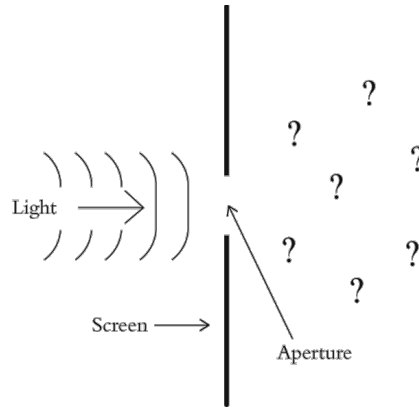
This section of the paper discusses a few solutions to the puzzle and argues that they are unsatisfactory or incomplete in some way. Here is an initial response to the challenge that ought to seem plausible. When faced with the challenge from inconsistent idealizations, one might observe that even if an apparatus contains an internal inconsistency, it will rarely be the case that *every* part of the apparatus is inconsistent with every other part. One might accordingly modify inferentialism in a way that meets the challenge: one might suggest that an apparatus is a representation of whatever target system its *noncontradictory* parts facilitate surrogative inferences about. The right kind of surrogative inference, on this view, is the kind that uses only the non-contradictory parts of a representation. In other words, the representational capacity of an internally inconsistent representational apparatus is constituted by the inferential capacity of its consistent sub-representations.<sup>4</sup> This would seem to resolve the problem posed by internally inconsistent idealizations.

This solution is unsatisfactory. One initial reason to reject it is that it may not generalize. There are some cases of inconsistency in a representational apparatus where there may not be any consistent sub-representations. Kirchhoff's diffraction theory, previously mentioned, might be one example (Saatsi and Vickers 2011).

But the real reason why this revised inferentialism should be rejected is that on inspection, it does not succeed at its appointed task. The view cannot account for the source of an inconsistent model's representational power. For even if an internally inconsistent representation has a number of consistent sub-representations, those sub-representations cannot explain why the *entire* inconsistent thing represents what it does. Let me illustrate by considering Kirchhoff's diffraction theory in more detail. It is a representation of a light-screen-aperture system, visualized below.

---

<sup>4</sup> This response bears a number of similarities to the “content-driven” means of explaining why internally inconsistent scientific theories are empirically successful (Norton 1987, Saatsi 2014). Also, this response bears a misleading resemblance in name only to the “partial structures” view of interpreting inconsistent scientific theories, defended by Bueno and French (2011), da Costa and French (2003), among others.



The theory contains the following content (some of which is obviously idealized):

- i. The Fresnel-Kirchhoff diffraction formula describes a light's post-aperture amplitude at some point P,  $U(P)$ :

$$U(P) = -\frac{iA}{2\lambda} \iint_A \frac{\exp(ik(r+s))}{rs} [\cos(n,r) - \cos(n,s)] dS$$

- ii. The light at the aperture behaves as if the screen were not there,
- iii. The light source emits a spherical field,
- iv. And the amplitude of light and its derivative are zero immediately behind the screen.

(Some purely mathematical parts of the model, i.e. the “Sommerfeld radiation condition”, are omitted.) To briefly describe the inconsistency in this theory, (ii)--(iv) and the omitted mathematical portions of the theory (the Helmholtz-Kirchhoff integral theorem and the Sommerfeld radiation condition) together derive (i), the Fresnel-Kirchhoff diffraction formula. But that formula in (i) is inconsistent with what (ii) and (iv) themselves say about the light-screen-aperture system, specifically  $U(P)$  and its derivative.<sup>5</sup>

The suggestion is supposed to be that the representational power of the entire inconsistent theory, and the facts about what that theory represents, are constituted by the representational and inferential power of proper subsets of  $\{(i)\dots(iv)\}$ . But consider what those sub-representations are actually like. They ascribe very different properties to the light-screen-aperture system than the conjunction of all four. That is, they represent that concrete target system as being very different

---

<sup>5</sup> See Saatsi and Vickers (2011). Their formulation of the Fresnel-Kirchhoff diffraction formula, which I follow here, is taken from Born and Wolf (1999, ch.6). For more discussion of the history of Fresnel's theory of diffraction, see Buchwald and Yeang (2016), who discuss, among other things, how Kirchoff's inconsistent theory remained more empirically successful than some later alternatives which were internally consistent.

ways. Or better yet: consider them in the inferentialist's preferred way, as tiny worlds that are surrogates for the concrete target system. Each of those sub-representations constitutes a very different “surrogate system” than the one that is constituted by conjoining all of (i)--(iv). Each system has distinct properties. Indeed, most of these different representations disagree with the others about whether some element in the target system has a particular property.

This is important because, to begin with, it seems false that each of these candidate surrogate systems is a *sub*-representation of the bigger, inconsistent thing they disagree with. But additionally, it's not clear how each of these differing representations is supposed to account for the representational power of the entire, inconsistent apparatus. For again, each of these “sub”-models and the entire, inconsistent model disagree about which surrogate facts there are. If they disagree with one another, then how are they also jointly supposed to generate some consistent representational power? Because of this problem, this solution is unable to account for there being some univocal fact about which target the model as a whole represents and how well it does so.

## 4.2 Hybrid Inferentialism?

A second kind of response that the inferentialist might make is to suggest that the admissible, representation-constituting inferences are those that are allowed by the parts of the model that are similar, isomorphic, homomorphic, etc., to the target system. While every inferentialist view is happy to agree that structural or qualitative features of a representational apparatus are what “allow for” or facilitate surrogative inference (Suárez 2004, p773), this suggestion goes one step further and hybridizes inferentialism with another view of scientific representation. On the view being suggested, it is a necessary condition on the surrogative inferences that constitute a model's representational capacity that they be performed not just with *any* structure of the model, but with those parts that are sufficiently similar, isomorphic, or what have you. Since nature doesn't contain any contradictions that models might thereby map onto, it will always be impermissible to use a contradiction in an inconsistent model as the basis of a surrogative inference—so, *prima facie*, the problem is solved.



While this solution might seem plausible, it will ultimately prove unsatisfying, especially by inferentialists' own lights. This suggestion reintroduces the problem that representation *tout court* should not be conflated with accurate representation (cf. Suárez 2003), and indeed, part of the reason inferentialism was supposed to be attractive in the first place is that it does not make this mistake. For these reasons, this suggested solution should be rejected by inferentialists outright. To recall these problems, the relations '...is isomorphic to...' and '...is similar to...' will only ever apply to the parts of a representational apparatus that *successfully* stand in the relevant relation with regard to the target. Agnes Bolinska notes, too, that any way of filling in how the structure of a model allows for surrogative inferences that trades on notions like isomorphism and similarity would "imply that a vehicle can be an epistemic representation of a target only if a user is able to draw inferences that are true of the target, yet Suárez explicitly states that the inferential account is meant to be an account of mere, rather than true or accurate, representation" (Bolinska 2013, p225). What this means is that if one has a representational apparatus with even a minor inaccuracy, a hybridized inferentialism won't have the resources to explain how it manages to be a representation at all. This familiar worry is especially pressing in cases of models that contain idealizations like dimensionless quantities, i.e., quantities in a model for which there is no corresponding physical parameter or quantity. It would be a hard bullet to bite indeed to suggest that no admissible representational surrogative inferences can be performed using parts of a model which are heavily idealized or dimensionless.

These sorts of problems are well-known, and defenders of other accounts of representation may well have strategies for responding to them. But the present point is that if there is any satisfying role for relations like similarity on an inferentialist account of representation, it must be that such relations facilitate successful representation without constituting it or being sufficient for it. Anjan Chakravartty, writing on this same point, argues that,

"Scientific representation is achieved only in circumstances in which agents know or have otherwise mastered the system of representation being used to encode information about whatever it is that is represented. That is why relations such as similarity cannot do the job on their own." (2010, p205)

The puzzle we're currently concerned with is what explains why only some, but not all, surrogative inferences performed using an internally inconsistent model are constitutive of its representational

capacity. As Chakravartty observes, one in general should not expect relationships of similarity to suffice by themselves for accounting for the intentionality involved in scientific representation. Hence even when it comes to our challenge regarding inconsistent idealizations, there is no reason to suspect that relationships like similarity will by themselves account for those intentional or surrogative features of representation deemed essential in any inferential picture. It is plausible that agents' practice, skills, or at least the intended goals of representation are also required for a full solution.

### 4.3 Cognitive Factors?

Let us now turn to agential factors, especially, cognitive factors. So far, we have considered solutions to the challenge from inconsistent idealizations that turn on something about the intrinsic features of a representational apparatus (e.g., consistent sub-representations). But as just mentioned, one might suggest that factors that concern agents' own cognitive states regarding scientific representations can provide more promising solutions.

First, consider a version of a view on which agents' *interpretations* of inconsistent models are what explain why explosive surrogative inferences don't help constitute those models' representational capacity. Gabriele Contessa defends a version of inferentialism on which "a model is a representation of a certain system in virtue of the fact that a user interprets the model in terms of the system" (2007, p67). Surrogative reasoning is still essential to scientific representation, but in fact surrogative reasoning is derivative on agential interpretation. What motivates this picture is the idea that without any interpretation in terms of a target system, the process of surrogative inference would just amount to tinkering around within a model for its own sake: it is interpretation that turns a model into a surrogate for a target, and hence a representation.

If one adopts a view like Contessa's, where the surrogative inferences a model licenses are due to an antecedent interpretation of the model, then there seems like there'd be an answer to the challenge posed by internally inconsistent models. Maybe this is the answer: surrogative inferences that exploit contradictions in a model aren't among those that are constitutive of a model's representational capacity, because there are no contradictory interpretations of inconsistent models.

This solution certainly sounds promising. But unfortunately, it will not succeed. The reason is due to what it means to have an interpretation of a model, on inferentialist views of representation. Remember how the interpretation-step of the surrogate reasoning process works: a user of the model interprets some surrogate fact as a fact about the target system. All this requires is that there is some interpretation or another— not necessarily a good or truth-preserving one—of parts of the model in terms of the target, and you can fruitfully complete the surrogate process. For just like other inferentialist views, Contessa’s interpretational view is supposed to be an account of representation *tout court*, without conflating it with successful or accurate representation (Contessa 2007, p63).

There are thus two problems with this interpretational view. First, interpretation does not avoid the challenge from inconsistent idealizations at all. For the only place that interpretation comes into the surrogate reasoning process is after some within-the-model facts have been ascertained. As soon as the model itself has a particular structure, all that an interpretation can do is “read off” that structure in terms of the target system. If there is a contradiction within a model, an interpretation can’t get rid of it, and it more importantly can’t disallow its use within the model. Second, an interpretational view actually makes the problem *worse* in some respects. Remember, an interpretation only requires *some* denotation of target-objects using model-objects. Indeed, as Agnes Bolinska (2013, p223) argues, one can merely stipulate by *fiat* some interpretation that connects the model to any arbitrary target. But this means that an interpretational inferentialism overgenerates representational capacity in much the same way that explosive inferences using an inconsistent model threaten to: just stipulate whatever interpretation you like! Perhaps a different definition of what interpretation requires can avoid this, but as-is, the standardly thin definition of interpretation on inferential views in fact exacerbates the overgeneration problem.

A second feature of agents’ cognitive factors that might help resolve the challenge from inconsistent idealizations is the extent to which an agent has know-how regarding the proper use of a model. In particular, Javier González de Prado Salas et al. (2017) argue for an inferentialist view of *idealization* (rather than representation more generally) that allows for degrees of cognitive commitment. For instance, when using idealizations, “one becomes committed to it just within the

scope of the relevant episode of hypothetical or approximative reasoning” (González de Prado Salas et al., 2017, p15).<sup>6</sup> This seems quite plausible: accepting an idealization in a model does not involve believing that it is true; it merely means utilizing its predictive capacity, then discharging soon after. This “temporary acceptance” picture of the epistemic utility of idealizations can be applied to present purposes. If an agent has the right know-how regarding the use of an inconsistent model, she will know how to use each half of a contradiction to perform some surrogative inference, then discharge it, thereby avoiding any contradiction and any explosive surrogative inferences.

While this view is plausible in the aforementioned respects, one must be careful to note that it cannot *solely* be agential know-how or cognitive states that rule out certain surrogative inferences as impermissible or not genuinely representational. For such a suggestion would turn inferentialism too much into an “agent-based” view of representation, i.e., one on which representation depends entirely on some agent or another’s occurrent cognitive states. Such a heavily agent-based solution would actually fail, in the end, to recover the kernel of inferentialism about representation. For by the lights of inferentialism about scientific representation (as opposed to, say, inferentialism about the epistemic utility of idealization), what it is for a model to represent its target just is for *it* or its structure to allow for surrogative inferences (cf. Suárez 2004, p771; Contessa 2007, p53; Kuorikoski and Ylikoski 2015, p3835). Of course, this will still involve some kind of reference to agents’ abilities and interests, but it is thinner than any reference to occurrent cognitive states. Contessa (2007, p52) makes this point explicit: “a user does not necessarily need to perform any *actual* piece of surrogative reasoning in order for the vehicle to be an epistemic representation of the target...she would have been able to perform one of those inferences if only the occasion had arisen.” In other words, for a full-blooded inferentialist view of representation, the representationmaking facts are supposed to be due to models’ own inferential potential, not due to any occurrent mental states like know-how or temporary commitment. Chalking up representation to something about agents’ know-how ends up being too much of a modification. Even if there weren’t any users who occurrently knew the skillset associated with adopting a model, it should still be the case that that model represents its target by facilitating certain inferences regarding it.

---

<sup>6</sup> Thanks to an anonymous referee for this suggestion.

So, while initially helpful, inferentialists must move on from an agent-based view like this one or at least clarify it in a way that is more amenable to the spirit of the view.

## 5 Surrogative Inferences and the Representation Game

### 5.1 Playing the Representation Game

Let me briefly take stock of where we have ventured so far. We have seen that inferentialist accounts of scientific representation face a challenge posed by internally inconsistent scientific models. Without a restriction on the kind of surrogative inferences that contribute to an apparatus's representational power, a restriction that rules out trivial, "explosive" inferences, we would have the implausible result that an internally inconsistent model represents absolutely everything (but merely trivially).

I wish to introduce an analogy. Many single-player and competitive games contain exploitable glitches in gameplay. One could list off many examples of games like this. But the most illustrative examples for our purposes are glitches in computer or video games that allow for the completion of certain tasks outside the normal bounds of gameplay. If, for instance, a task that a video game requires of the player is that the player avatar retrieve an item from within a particular cage, it might be the case that there are a few permitted ways to do it, all of which involve actually opening the cage. But a common glitch in these kinds of scenarios is that the player can manipulate game mechanics in such a way that her avatar is pushed *through* the normally-impenetrable walls of the cage. By exploiting these kinds of behaviors—being pushed through walls, falling through floors, etc.—players can accomplish the same tasks they are asked to accomplish normally but in much shorter timespans (e.g., in "speedrunning").

Now, consider the fact that if a player retrieves the caged item by exploiting a glitch, there is at least some sense in which she hasn't done anything wrong—after all, the game doesn't crash. Moreover, glitches are in fact caused by the way that certain parts of actual gameplay are coded (e.g., how one sprite's "hitbox" behaves, or in other cases, the developers neglecting to put a ceiling

in some rooms). In other words, the game indeed does allow for such events to occur. However, even though the game in fact allows for these events to occur, there is a clear sense in which, if a player uses the exploits, she fails to successfully *play* the game. Retrieve an item by opening its cage in any of the intended ways? That constitutes actual gameplay. Retrieve it by glitching through the walls? That is allowed, but there is a clear sense in which it does not really count as having played the game. Even though both kinds of actions are “allowed” by the game, the first but not the second constitutes gameplay because it uses the game’s elements to serve their intended functions.<sup>7</sup>

Internally inconsistent scientific representations are like games with exploitable glitches. The behaviors and properties of their constituents can be manipulated, allowing for tasks to be completed that could *also* be completed by obeying those behaviors’ proper role and scope. Recall Kirchhoff’s theory of diffraction and Bohr’s model of the atom. Kirchhoff’s theory is remarkably accurate regarding changes in the amplitude of light, Bohr’s regarding emission and absorption spectra, yet both contain contradictions. The simple but important observation I want to make is that despite this obvious problem, there is nothing objectionable about using the Fresnel-Kirchhoff formula to calculate the post-aperture amplitude  $U(P)$  or using a graphic of electron orbitals in Bohr’s model to teach students true things about molecular bonding. And the reason seems to be that doing so would constitute a use of those features of the models that they were intended for. On the other hand, if one chooses to “speedrun” the model by generating a contradiction, it is quite clear that generating a contradiction and exploiting some features of classical logic cannot be the intended function of having represented, e.g., electrons or light a particular way. This seems true even though it would not be in violation of any logical principles to exploit the contradictions inherent in inconsistent models. In other words, there are *non-logical* norms on what constitutes representational “gameplay” that are exerted by features about the model, or its context or aims.

---

<sup>7</sup> Zamora Bonilla (2006) defends an idea of the aims of scientific research as a persuasion game, drawing on the analogous notion from the philosophy of language. My approach here is aimed at understanding scientific *representation*, not entire research programs, and also does not rely on the idea of a “persuasion game” in philosophy of language. Nevertheless, because his account of the aims of scientific research is both inferentialist in nature and relies on the idea of gameplay, there are similarities between his work and mine.

I want to be careful here before proceeding further to clarify that my appeal to “gameplay” is a mere analogy. Indeed, while some other recent developments in philosophy of science have appealed to games, there is at least one prominent case where inferentialists should not and cannot avail themselves of a deeper analysis of representation in terms of gameplay.<sup>8</sup>

Prominently, some recent work on scientific representation claims that idealized models either literally are fictions or are analogous in important respects to fictions (Frigg 2010, Toon 2010, 2012, Levy 2015, among many others). Of particular note are Adam Toon (2010, 2012) and Arnon Levy (2012, 2015), who hold that scientific representation is to represent targets in a fictionalized way. Their views are noteworthy because Toon in particular develops an analogy with Kendall Walton’s (1990) idea of “games of make-believe.” Like my account, Toon’s account of representation makes use of the idea of rules of gameplay (specifically, Toon’s view has it that idealized models involve games of make-believe about concrete targets, e.g., imagining chemical bonds as the wooden sticks in a ball-and-stick molecule).

However, Toon and Levy defend the view that scientific representation is (as they say) *direct*. It’s not the case that representation involves any relation at all between a distinct model system and a target system, as inferentialists and others would have it; instead, it merely involves imagining false things about the target. This difference is a deep one. Levy (2015) himself notes that direct representation leaves unanswered the question of how, if at all, model users are to infer or learn anything about the target from such imaginings. And Frigg and Nguyen (2016) observe that imaginings about a target don’t license *any* inferences about what the target is actually like. Even within Walton’s notion of “make-believe”, they show, nothing accounts for how surrogative reasoning can ever get off the ground, let alone be informative. In other words, direct representational views like Toon’s “games of make-believe” cannot easily and obviously account for how surrogative reasoning ever even happens. Inferentialists, for whom surrogative reasoning is literally constitutive of scientific representation, should thus be quite wary of taking

---

<sup>8</sup> An anonymous referee helpful points out that initial area of similarity to previous gameplay-based analogies is to Jaakko Hintikka’s (1973) game-theoretic semantics for natural language and first-order logic, which Jesús Zamora Bonilla (2003) has utilized for understanding properties of scientific theories, like testability.<sup>8</sup> This extant literature is primarily concerned with giving a full analysis of certain semantic concepts in science in terms of certain kinds of gameplay in natural language, while my aim is to identify how non-logical norms on surrogative inferences may arise. Nevertheless, there may be a way to connect these issues.

representation to *literally* be a game of make-believe, or indeed, any direct view of scientific representation. There may yet be a solution to this problem in the offing, but my aim throughout this paper has been to defend inferentialism or modify it in a friendly way, and the simplest way to do so is without adopting a view such as Toon's.

Of course, inferentialists are not precluded from utilizing analogies to gameplay or fiction in other ways. Frigg and Nguyen develop a broadly inferentialist picture where representational *constraints* are “naturally analyzed as being involved in a game of make-believe” (2016b, p237). Suárez (2009) defends an inferentialist account of representation under which idealizations, viewed as fictions, provide background rules for expediency in surrogative inference. These other ways in which inferentialism can make use of analogies to gameplay or fiction are not motivated by worries about trivialization or inconsistent models, as my paper has been. And I have no objection to the possibility that these other concerns might ultimately motivate an inferentialism integrated with my own. Unfortunately, there isn't space here to engage in more detail with these views, but it is again worth noting that as far as my own account goes, “gameplay” is merely an analogy. What is most central to my view, to reiterate the point introduced before, are the ways in which the behaviors and intended functions of particular parts of models can exert non-logical norms on how surrogative inferences can proceed.

#### **5.4 Explicating the View**

The only task remaining in this paper is to further clarify this key point. How can the behaviors and intended functions of particular parts of models exert non-logical norms on which surrogative inferences are permissible and constitutive of representational capacity? First, it is a datum that the components of a model have certain behaviors that comprise the ways in which scientific representations take on “a life of their own” (cf. Hughes 1997, pS331). So, before introducing how the intended functions of a model's parts exert norms on which kinds of inferences are the legitimately representational ones, I wish to illustrate how the behavior of the internal world of a model itself does so. Consider how, in the world of a model, a quantity might be quantized rather than continuous, or its rate of change might be continuously differentiable; some variable might be designated to evolve nonlinearly, one variable's growth might be causally dependent on another,



and so on. To see how inferential norms arise from these, remember that surrogative inferences are supposed to start by determining what happens in the world of a representation, and consider a simple example using the ideal gas law  $PV = nRT$ . Imagine that a student uses the ideal gas law to calculate the difference in temperature of a particular sample of gas as the gas's volume is decreased, and another student says: "Oh, no, you've done that wrong: your value for the final temperature needs to be lower than what you've calculated, since all gases in fact lose temperature over time."

It seems clear that it is this second student who is actually in error. By suggesting that the first student's use of the ideal gas law's mathematical model needs to incorporate the slight losses in temperature of observed gases, she misunderstands how that apparatus is to be used. She gets its world wrong. Correcting the ideal value when we use ' $PV=nRT$ ' obviously does not generate a contradiction, and in fact in many contexts it might be the sort of operation that is eventually required. So—importantly—there are no purely logical norms against correcting from the ideal temperature. But it is clearly a mistake to suggest that *the model itself* licenses the use of its own concept of temperature in a non-ideal way. The rules of the game of ' $PV=nRT$ '—its non-logical norms, in other words—declare the only permissible inferences within the model system to be those that accord with the idealizations of the kinetic theory of gases. So, if you were to infer within the model that a gas's temperature falls off, clearly that would not be a use that constitutes the content with which that the model was intended to represent the world.<sup>9</sup>

This is importantly different from the kind of view considered earlier, which was due to González de Prado Salas et al. (2017). By their lights, the sort of acceptance-then-discharging of the idealizations that this second student engages in is perfectly in line with cognitive states that would constitute a more general knowledge-how of using the model. But in the vignette given above, the second student is clearly in error because she corrects for a part of the model that violates the model's own structure and gets its "world" wrong. But she is in error not just because she gets the world of the model wrong. It is a mistake, moreover, because regardless of cognitive states or her

---

<sup>9</sup> "...the representation has an internal dynamic whose effects we can examine [...] The same result appears whether we use the mathematical or the material model. The internal dynamic of the mathematical model is supplied by a mixture of geometry and algebra, that of the material model by natural processes... The internal dynamic of a computer simulation of the phenomenon would be something else again." (Hughes 1997, pS332).

correct factual knowledge regarding the intended physical targets, she is not using the model in its intended ways. One might say that the aim of the ideal gas law (and the kinetic theory of gases more generally) is to represent gases in an ideal way to fulfill certain functions, say, calculational tractability, or historically, for coherence with phenomenological thermodynamics. The fact that the so-called world of the model has certain features is thus tied to those functions. Failing to respect the model's own world, even if it achieves accuracy *simpliciter* (like in the preceding vignette), can nevertheless mean that one is not using the model in its intended function.

Hence, in identifying how non-logical norms on surrogative inferences are generated, the intended scope and function of the model's properties and idealizations play the most important role. To see this point most clearly, it can be helpful to rely on a distinction first introduced by Mary Hesse.<sup>10</sup> Speaking of billiard-ball models of gas collision, she writes:

“Let us call those properties we know belong to billiard balls and not to molecules the *negative* analogy of the model. Motion and impact, on the other hand, are just the properties of billiard balls that we do want to ascribe to molecules in our model, and these we can call the *positive* analogy.” (1966, p8)

The point is that some but not all parts of models are intended by the initial modeler to have a representational function. One can endorse this general point without adopting Hesse's distinction or even adopting wholesale the view that models are primarily analogies. It would be an obviously illegitimate use of an ideal billiard-ball model of gas collision, say, to use features of the model which are not intended in the first place to be representational in a representational way. Such illegitimate uses wouldn't be informative at all regarding what the model represents, or how well it does so, etc. Inferentialists can easily adopt this idea. The legitimately representational uses of a model are limited to surrogative inferences performed using parts of the model that were either intended in the first place to be useful surrogatively, or those which might still (through further investigation) have some surrogative import.<sup>11</sup> (Hesse calls this last kind “neutral”.)

---

<sup>10</sup> Thanks to Darrell Rowbottom for the helpful pointer about Hesse's distinction, as well as some discussion on this general point.

<sup>11</sup> Darrell Rowbottom, writing on Hesse's distinction as well, notes that “the *positive analogy* should be recast as involving the set of properties that the modeller expects to do some work (in furthering predictive power, understanding, or what have you)” (2019, p42). Additionally, it is worth noting that the so-called “neutral” analogy, can help explain what justifies the application of models constructed for one domain into another, in cases of “model transfer”.

When it comes to internally inconsistent models, the way in which the intended surrogative or representational aim of the parts of a model allow or disallow certain uses becomes even clearer. I have said that generating a contradiction and exploiting some features of classical logic cannot be the intended function of having represented, e.g., electrons or light a particular way. Indeed, even if those involved in constructing an internally inconsistent model were aware of the inconsistency, it is surely not the case that they intended the contradiction in a model to be exploited for nefarious, explosive purposes.<sup>12</sup>

So, we have a solution at last to our puzzle about inconsistent idealizations. Why is it that inconsistent models don't universally represent? Moreover, why is it the case that possible uses of inconsistent models that might exploit a contradiction don't contribute anything toward its representational capacity, i.e., what it represents, or how well it does so? The answer is that the only legitimately representational uses of a model are those that cohere with the intended function of inferring something about the world, or put another way, inferring some information about some parts of the target in a surrogative capacity.<sup>13</sup> The intended function of representing electrons, or whatever else, a particular way, must have been to infer information about the world, not to cheat classical logic. So, any usage of the representation that fails to even aim at that function—such as one that only aims to cheat classical logic—cannot be representation-constituting.

The intended aims of a model during model-building also explain why the application conditions of some parts of a model are limited. Let us revisit Bohr's model again, to illustrate how these informational aims determine which surrogative inferences are legitimate. Peter Vickers writes that "Bohr assumes that classical electrodynamics can be trusted, more or less, in its traditional domain of application. But the obvious flipside of this is that Bohr is assuming that classical electrodynamics does *not* apply to electrons strongly bound to atoms (with relatively low quantum

---

<sup>12</sup> And even *if* Bohr, or Kirchhoff, or whoever else, turned out secretly to have been a dialetheist, this would *still* not license exploiting a contradiction and using the principle of explosion, since most dialetheists reject the validity of explosion.

<sup>13</sup> Or, as Agnes Bolinska (2013) puts it, the intended function of interacting with or interpreting a model is to achieve *informativeness* about the target through faithful representation. See Bolinska's paper for a more thorough discussion of varieties of interpretation in scientific representation, and what it means to have faithful representation as an aim.

number  $n$ )” (2013, p50). In other words, while it is trivial to say that the intended function of any model is to provide information, the classical portions of the Bohr model aim at informativeness about a particular domain of application. The fact that the model contains classical parts *at all* means that, if some user wanted to, she could exploit them to generate a contradiction with some relevant nonclassical parts. (This is, again, just like the way in which the code of a glitched game in some sense genuinely “allows” for their glitches to occur.) But doing so means using the model’s elements outside of their intended domain. Generating contradictions and exploiting some principles of classical logic are functions for which they could not possibly have been intended.

This example also wards off an objection that, as we saw in previous sections, afflicts many other accounts of scientific representation: conflating accurate representation with representation *tout court*. Even though Bohr’s model was inaccurate regarding, e.g., the spin of electrons in the ground state, it should still clearly count as a model of the atom. But as we saw, solutions that say that the legitimately representational surrogative inferences are only those that use similar-or-isomorphic parts of a model runs afoul of this datum. On the other hand, on my analysis, using a within-the-model fact about “Bohr electrons” to surrogatively infer false conclusions about actual electrons still counts as legitimately representational. Clearly, that is a use of Bohr electrons that fulfills their intended function, i.e., information about actual electrons. As long as one uses the parts of a model to at least *try* to infer information about the corresponding worldly target, one can represent that target without successfully gaining information. To revisit the gameplay analogy, it is possible to successfully play a game without winning it.

Before closing, one might worry whether my view about function-cohering uses of a model can fully generalize beyond the example of Bohr’s atom that I have used so far. To be clear, Bohr’s atom is merely a helpful example since its derivational context is quite explicit. But even without this degree of explicitness, the broader point about legitimately representational uses of a model can still be made. And that broader point, which is fully general, is that the function of a model and its parts is to represent the world and provide information about a worldly target. This intended function provides an accordingly general explanation of why “explosive”, contradiction-utilizing surrogative inferences do not count as representational. The only legitimate surrogative inferences are those that use a model’s parts for their intended function. And nature cannot contain any

contradictions, complaints from dialetheists be damned. So, it is literally impossible for a use of a model that generates a contradiction to count as a function-fulfilling use, much less a function-aimed use. Again, generating a contradiction cannot have been the intended function of having represented electrons a particular way. (Barring, of course, the sudden appearance *per impossibile* of a contradiction in the wild.) Even with inconsistent models, the only moves in “gameplay” that genuinely constitute a model’s representational capacity are those which cohere with its parts’ intended function.

## **6 Conclusion**

I have argued that inferentialist views of scientific representation, left unamended or unprecisified, face a serious challenge. Because many representational apparatuses in science are idealized in internally inconsistent ways, they appear to support an unlimited number of valid surrogative inferences regarding their target, and any target. Clearly, however, inconsistent models do not represent their targets with unlimited inferential capacity, nor do they represent absolutely everything. So, there needs to be an explanation of why the possible contradiction-based surrogative inferences in an internally inconsistent model do not contribute to its representational capacity. I have argued that the inbuilt behavior of the “world of the model” exerts norms on how its parts may be used in surrogative inferences. And even more than the behavior of the world of the model, I have argued that a model’s parts’ intended function draws principled boundaries on how the properties of the model can be used. For even in cases of internally contradictory models, the intended function of a model’s representational content is to facilitate inferences about the world, not to facilitate exploitative behavior. To sum up, by developing the ways that the aims of scientific representation can exert non-logical norms on how they are used, I have shown how inferentialists can avoid the challenge from inconsistent idealizations, as well as further clarified exactly what the components of surrogative inference should be on an inferentialist view of scientific representation.

## **Acknowledgments**

For helpful comments on previous versions of this paper, thanks to Harjit Bhogal, Jaakko Kuorikoski, Darrell Rowbottom, Tom Wysocki, audiences at the University of Virginia, Models

and Simulations 7 in Barcelona, and the APA Eastern 2019. Special thanks go out to two anonymous referees for this journal, Manuel Barrantes, Anjan Chakravartty, Paul Humphreys, and James Nguyen.

## Works Cited

- Bolinska, A. (2013). Epistemic representation, informativeness and the aim of faithful representation. *Synthese* 190 (2):219-234.
- Born, M. and Wolf, E. (1999). *Principles of Optics*, 7<sup>th</sup> (expanded) edition. Cambridge University Press.
- Buchwald, J., and Yeang, C. (2016). Kirchoff's theory for optical diffraction, its predecessor and subsequent development: the resilience of an inconsistent theory. *Archive for History of the Exact Sciences* 70: 463-511.
- Bueno, O. (2014). Computer Simulations: An Inferential Conception. *The Monist* 97(3):378-398.
- Bueno, O., & Colyvan, M. (2011). An Inferential Conception of the Application of Mathematics. *Noûs* 45 (2):345-374.
- Bueno, O., and French, S. (2011). How Theories Represent. *British Journal for the Philosophy of Science* 62 (4):857-894.
- Chakravartty, A. (2010). Informational versus functional theories of scientific representation. *Synthese* 172 (2):197-213.
- Contessa, G. (2007). Scientific representation, interpretation, and surrogate reasoning. *Philosophy of Science* 74 (1):48-68.
- da Costa, N.C.A. & French, S. (2003). *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning*. Oxford University Press.
- de Donato Rodríguez, X., & Zamora Bonilla, J., (2009). Credibility, Idealisation, and Model Building: An Inferential Approach. *Erkenntnis* 70 (1):101-118.

Frigg, R. (2010). Models and fiction. *Synthese* 172 (2):251-268.

Frigg, R., & Nguyen, J. (2016). Models and Representation. In L. Magnani & T. Bertolotti (eds.), *The Springer Handbook of Model-Based Science*. Springer. 49-102.

Frigg, R., & Nguyen, J. (2016b). The fiction view of models reloaded. *The Monist* 99 (3):225-242.

González de Prado Salas, J., de Donato Rodríguez, X. & Zamora Bonilla, J. (2017). Inferentialism, degrees of commitment, and ampliative reasoning. *Synthese* online first: 1-19. DOI 10.1007/s11229-017-1579-5.

González de Prado Salas, J., Suárez, M. & Zamora-Bonilla, J. (2018). Special issue: Inferentialism in philosophy of science and in epistemology—introduction. *Synthese* online first: 1-3. DOI 10.1007/s11229-018-02009-4.

Hesse, M. (1966). *Models and Analogies in Science*. University of Notre Dame Press.

Hintikka, J. (1973). *Logic, Language-Games and Information*. Oxford, Clarendon Press.

Hughes, R.I.G. (1997). Models and representation. *Philosophy of Science* 64 (4): S325-336.

Kuorikoski, J., & Ylikoski, P. (2015). External representations and scientific understanding. *Synthese* 192 (12):3817-3837.

Levy, A. (2012). Models, Fictions, and Realism: Two Packages. *Philosophy of Science* 79 (5):738-748.

Levy, A. (2015). Modeling without models. *Philosophical Studies* 172 (3):781-798.



Norton, J. (1987). The logical inconsistency of the old quantum theory of Black body radiation. *Philosophy of Science* 54 (3):327-350.

Rowbottom, D. (2019). *The Instrument of Science: Scientific Anti-Realism Revitalized*. Routledge.

Saatsi, J., & Vickers, P. (2011). Miraculous Success? Inconsistency and Untruth in Kirchhoff's Diffraction Theory. *British Journal for the Philosophy of Science* 62 (1):29-46.

Saatsi, J. (2014). Inconsistency and scientific realism. *Synthese* 191 (13):2941-2955.

Suárez, M. (2003). Scientific representation: Against similarity and isomorphism. *International Studies in the Philosophy of Science* 17 (3):225-244.

Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of Science* 71 (5):767-779.

Suárez, M. (2009). Scientific fictions as rules of inference. In *Fictions in Science: Philosophical Essays on Modeling and Idealization*. Routledge. pp. 158--178.

Suárez, M. (2010). Scientific representation. *Philosophy Compass* 5 (1):91-101.

Suárez, M. (2015). Deflationary representation, inference, and practice. *Studies in History and Philosophy of Science Part A* 49:36-47.

Sugden, R. (2000). Credible worlds: the status of theoretical models in economics. *Journal of Economic Methodology* 7 (1):1-31.

Sugden, R. (2009). Credible Worlds, Capacities and Mechanisms. *Erkenntnis* 70 (1):3-27.

Swoyer, C. (1991). Structural representation and surrogative reasoning. *Synthese* 87 (3):449 - 508.

Toon, A. (2010). Models as make-believe. In R. Frigg & M. Hunter (eds.), *Beyond Mimesis and Convention: Representation in Art and Science*. Boston Studies in Philosophy of Science.

Toon, A. (2012). *Models as Make-Believe: Imagination, Fiction, and Scientific Representation*. Palgrave-Macmillan.

Vickers, P. (2013). *Understanding Inconsistent Science*. Oxford University Press.

Walton, K. (1990). *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Harvard University Press.

Zamora Bonilla, J. (2003). Meaning and testability in the structuralist theory of science. *Erkenntnis* 59 (1):47 - 76.

Zamora Bonilla, J. (2006). Science as a persuasion game: An inferentialist approach. *Episteme* 2(3): 189–201.