REPRESENTATION IN SCIENCE

Oxford Handbook in Philosophy of Science

Mauricio Suárez (Institute of Philosophy, London University, and Complutense University, Madrid)

Abstract: This article provides a state of the art review of the philosophical literature on scientific representation. It first argues that the topic emerges historically mainly out of what may be called the modelling tradition. It then introduces a number of helpful analytical distinctions, and goes on to divide contemporary approaches to scientific representation into two distinct kinds, substantive and deflationary. Analogies with related discussions of artistic representation in aesthetics, and of the nature of truth in metaphysics are pursued. It is finally urged that the most promising approaches - and the ones most likely to feature prominently in future developments - are deflationary. In particular, a defence is provided of a genuinely inferential conception of representation.

Keywords: Scientific Representation, Modelling, Deflationism, Inferential Conception

Contents:

1. Historical Introduction

Scientific representation is a latecomer in philosophy of science, receiving considerable attention only in the last fifteen years or so. In many ways this is surprising since the notion of representation has long been central to philosophical endeavours in the philosophies of language, mind, and art. There are a number of historical reasons for the earlier neglect. At least initially this may be due to the logical empiricists' stronghold upon the field and their mistrust of notions of correspondence and direct reference – it was perhaps assumed that representation was one of those notions. In addition, in what has come to be known as the 'received view' (the name stuck even though it was only 'received' in the 1970s) scientific knowledge is articulated and embodied entirely within scientific theories, conceived as linguistic or propositional entities. Models play only a heuristic role, and the relation between scientific claims and theories on the one hand, and the real world systems that they are putatively about, on the other, is therefore naturally one of description rather than representation (Bailer-Jones, 2009). Although the distinction is not entirely sharp, a critical difference between description and representation concerns the applicability of semantic notions such as truth, which are built into descriptions but seem prima facie ill-suited for representations (as has been emphasized by e.g. Ronald Giere, 1988, Ch. 4). In focusing on the role that language plays in science, the logical empiricists and their successors may thus have implicitly privileged theoretical description.

The analysis of the logical structure of scientific theories remained a central concern for philosophers of science during the post war years, to the detriment of any thorough attention to the role that models, model-building and other genuinely representational entities and activities play in science. The rejection of the 'received' or syntactic view in favour of a 'semantic' conception in the 1980s did not in the first instance improve things much, since the focus continued to be the analysis of scientific *theory*. But the semantic view arguably contained the seeds of a deeper change of outlook since the central claim that theories could be understood as collections of models surreptitiously shifted attention from description onto representation. In particular Bas Van Fraassen's version of the semantic view in terms of state spaces, and Ronald Giere's in terms of cognitive models both emphasized how theories are tools for representation rather than description. These authors came to the conclusion that the linguistic analysis of scientific theory was of limited interest and emphasized instead the representational roles of models (Van Fraassen, 1980, Ch. 3; Giere, 1988, Ch. 3) There is no doubt that the development of this non-linguistic version of the semantic view was an essential step in the upsurge of representation in philosophy of science.

Nevertheless, the most important historical route to the notion of representation – and also the main source of current interest in it within the philosophy of science – is contributed by what we may call the modelling tradition, or the 'modelling attitude' (see Suárez, Forthcoming). This is the historical series of attempts by both philosophers and practicing scientists to understand and come to terms with model building, analogical reasoning, and the role that images, metaphors and diagrams play in modelling. The tradition has an interesting history too, beginning much further back in the works of philosophically informed scientists in the second half of the 19th century, such as William Thomson, James Clerk Maxwell, Heinrich Hertz, and Ludwig Boltzmann. Boltzmann's widely read article in *Encyclopedia Britannica*, in particular, signals the belle époque of this 'modelling attitude' which has done nothing but continue to flourish and inform much scientific practice during the 20th century (in spite of philosophical detractors such as Pierre Duhem, who famously disparaged against it in Duhem, 1906). Within the philosophy of science, it was mainly opponents to logical empiricist reconstructions of knowledge (both Carnapian and post-Carnapian) who pursued this modelling tradition, and continued to emphasize the essential role of models, model-building, and analogical reasoning in the sciences. Thus Norman Campbell's masterly *Physics: The Elements* (Campbell, 1920) had considerable influence in advancing the case for modelling amongst mainly British scholars such as Max Black (1954) and Mary Hesse (1962). Rom Harré (1960) was also influenced by Stephen Toulmin's (1960) vindication of theories as maps.

Within the sciences the modelling attitude has arguably been the prevalent methodology for acquiring predictive knowledge and control of natural and social systems ever since. Within philosophy of science, however, it has enjoyed more varied fortunes: After a peak of interest in the 1960s, the focus on models and representation waned considerably again, only to fully re-emerge in the late 1990's around what is known as the 'mediating models' movement. This was a movement of scholars based at the London School of Economics, the Tinbergen Institute and the Wissenschaftkolleg in Berlin, who developed and advanced a case for models and their role in scientific inquiry during the 1990s. The mediating models movement did not just look back in order to vindicate Mary Hesse's work in the 1960s but also proposed a view entirely of its own, according to which models are autonomous entities that mediate between theory and the world. Models are neither simply inductive generalizations of data, nor are they merely elementary structures of theory. Rather they are independent entities, very much endowed with a life of their own, and playing out a variety of roles in inquiry, which prominently include 'filling in' theoretical descriptions for their concrete application (Morrison and Morgan, 1999). The view lends itself to a certain sort of instrumentalism about theories in the building of models (Cartwright et al., 1994; Suárez, 1997) which paves the way for an understanding of models as representations.

The contemporary discussions of representation thus emerge from two somewhat distinct currents of thought: the semantic approach, and the mediating models movement. Some of the defenders of mediating models reject the semantic view but mainly on account of its construal of what models are (under the constraint that models are whatever provides identity conditions for theories); but it is nonetheless the case that scholars working in either movement emphasize the view that models are genuinely representational. [1] One of the most significant

---

[1] The notable exception maybe 'structuralism', which holds onto a set-theoretical version of the semantic conception arguably in conjunction with a kind of non-representationalism regarding scientific theory (Balzer et al., 1989).

pioneering papers (Hughes, 1997) is in fact the result of exposure to both movements. Hughes was already a leading defender of the semantic view (which he had successfully applied to problems in the philosophical foundations of quantum theory in Hughes, 1989) when he went on to become a prominent contributor to the mediating models movement (Hughes 1999). It is important to bear this dual heritage in mind since it goes some way towards explaining some of the inner tensions and open disagreements that one finds nowadays in this area.

2. Elements of Representation

There are many different types of representations in the sciences, in areas as diverse as engineering, mathematical physics, evolutionary biology, physical chemistry, economics. Modelling techniques in these areas also vary greatly as do the typical means for a successful application of a model. This is prima facie a thorny issue for a theory of representation, which must provide some account of what all these representations have in common. Nevertheless we may just consider one particular model, for our purposes, particularly since it has been widely discussed in the literature as a paradigmatic example, namely the '*billiard ball* model of gases'.

2.1. Sources and Targets

The billiard ball model is a central analogy in the kinetic theory of gases developed in the second half of the 19th century (Brush, 2003). Perhaps its first appearance – certainly the most celebrated one – in the philosophical literature occurs in Mary Hesse's work (Hesse, 1962, pp. 8ff) where the model is employed to distinguish what Hesse calls the 'negative', 'positive' and 'neutral' analogies: [2]

> "When we take a collection of billiard balls in random motion as a model for a gas, we are not asserting that billiard balls are in all respects like gas particles, for billiard balls are red or white, and hard and shiny, and we are not intending to suggest that gas molecules have these properties. We are in fact saying that gas molecules are *analogous* to billiard balls, and the relation of analogy means that there are some properties of billiard balls which are not found in molecules. 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* […] There will generally be some

[2] Hesse (1966, pp. 8ff.) traces the model back to Campbell (1920) although it is significant that the term "billiard ball model" does not appear there once – and in fact neither do billiard balls, but only more generic "elastic balls of finite diameter". The first explicit appearance of the full billiard balls analogy that I am aware of occurs in Sir James Jean's influential textbook (Jeans, 1940, p. 12ff). This is a significant point in historical scholarship, since the billiard ball analogy itself cannot have played the heuristic role that Hesse ascribes to it in the development of the theory; but such historical considerations need not detain us here.

properties of the model about which we do not yet know whether they are positive or negative analogies […] Let us call this third set of properties the *neutral analogy*."

In order to unravel the implications of this quote for representation, we need to first draw a number of distinctions. Let us refer to the system of billiard balls as the *source*, and the system of gas molecules as the *target*. We say that billiard balls represent gas molecules if and only if the system of billiard balls is a representational source for the target system of gas molecules. [3] The extensions of 'source' and 'target' are then picked out implicitly by this claim, i.e. any pair of objects about which this claim is true is a <source, target> pair. We can then list the properties of the source object as $\{P^s_1, P^s_2, \ldots, P^s_i, \ldots, P^s_j, \ldots, P^s_n\}$ and those of the target object as $\{P^t_1, P^t_2, \ldots, P^t_i, \ldots, P^t_j, \ldots, P^t_n\}$. The claim then is that some of these properties are identical: $P^s_1 = P^t_1$, $P^s_2 = P^t_2$, …, $P^s_i = P^t_i$. Hence $\{P^t_1, P^t_2, \ldots, P^t_i\}$ constitute the positive analogy. What are we say about the remaining properties? On Hesse's account we minimally know that some properties of the source system of billiard balls are not at all present in the system of gas molecules; thus e.g. $\{P^s_{i+1}, \ldots, P^s_j\}$ are not identical to any of the properties of gas molecules $\{P^t_1, P^t_2, \ldots, P^t_i, \ldots, P^t_j, \ldots, P^t_n\}$. These properties of billiard balls constitute then the negative analogy; while the remaining properties $\{P^s_{j+1}, \ldots, P^s_n\}$ constitute the neutral analogy (since we do not know whether they are in the positive or negative analogies). However, this is a purely epistemic criterion and we may suppose that all properties of billiard balls are objectively in either the positive or negative analogy: They are either really shared by both balls and molecules, or they are not, regardless of how accurate our knowledge of these facts is.

However, note that Hesse frames her notion of analogy in terms of the properties of the source: the criterion is that some of these properties obtain in the target and some do not. This criterion is symmetrical as long as analogy is understood as the sharing of identical properties; but even under this assumption, it does not provide some important information regarding the relation between the properties involved in the negative analogy. For it could be that that properties are absent altogether in the target, or that they are explicitly denied in the target. And this would amount to a major difference. (In other words it could be that there is some $P^t_{j+x} \in \{P^t_j, \ldots, P^t_n\}$, such that $P^t_{j+x} = \neg P^s_{j+w}$, for some $P^s_{j+w} \in \{P^s_j, \ldots, P^s_n\}$). For there is a difference between saying that gas molecules are unlike billiard balls in that they are soft (the property applies properly, but it is plainly contradicted), and saying that they are unlike billiard balls in that they are neither hard nor soft (the property just does not apply and there is no genuine contradiction). The latter statement seems more appropriate in the case of hardness / softness. But it need not always be more appropriate to deny the application of a property for the negative analogy. For example when it comes to e.g. elasticity, everything is quite different: billiard balls are after all only *imperfectly* elastic (Jeans, 1940, pp. 15-16) Thus there are different ways in which properties can be objectively in the

[3] The claim may be generic for any pair of types of such systems, or particular for one particular system of billiard balls with respect to a particular set of gas molecules enclosed in a container.

negative analogy. Billiard balls are hard and shiny and these properties are simply inapplicable to gas molecules. They are both in the negative analogy 'by absence'. But the ball's imperfect elasticity is certainly applicable – even though denied in the molecules, which are presumed to be completely elastic. The inelastic character of the billiard balls is in the negative analogy 'by denial'. As we shall see, the difference between absence and denial turns out to be of some significance for discussions of scientific representation.

2.2. Means and Constituents

Another important distinction that the example brings into relief is one between means and constituents of a representation (Suárez, 2003). The billiard balls provide a representation of the molecules in the gas, and as such share a number of properties with them. So one can reason from the properties of billiard balls in the model to the properties of molecules. The positive analogy therefore provides the material *means* that allow representations to do their work. But it would go beyond this minimal statement regarding the function of the positive analogy to assert that the positive analogy *constitutes* the representational relation. This is not merely a statement regarding the functional grounds of the representation, but rather the nature, essence or definition of the representation. There is a world of a difference between stating the positive analogy as what allows the representation to do its work (i.e. the reason why the representation is useful, accurate, predictive, explanatory, relevant, etc) and stating it as what the representation is essentially (i.e. the very constituent representational relation, its defining condition). We may summarize them as follows:

Means: R is the means (at a particular time and context) of the representation of some target t by some source s if: i) R (s, t); and ii) some user of the representation employs R (at that particular time and in that context) in order to draw inferences about t from s.

Constituent: R constitutes (at all times and in every context) the representation of some target t by some source s if and only if: i) R (s, t) and ii) for any source-target pair (S, T): S represents T if and only if R (S, T).

It is important to note both the inverted order of the quantifiers and the temporal index in the definition of the means. In other words, the constituent of a representation is an essential and perdurable relation between sources and targets, while the means are those relations that at any given time allow representation-users to draw inferences from the source about the target. It turns out to be an interesting question whether these coincide in general, or whether there is any need for a constituent at all. Note that this issue is independent of whether or not there are objective positive and negative analogies between sources and targets. Hence a fundamental question for theories of representation to determine is whether Hesse's positive analogy belongs to the constituents of representations (in which case it is essential to the representation itself) or to their means (in which case it is not so essential). The answer ultimately depends upon whether representation in general is substantial or deflationary.

3. Theories of Representation

A theory of scientific representation will aim to provide some philosophical insight into what all representations have in common, as well as what makes representations scientific. There are at least two different ways to go about providing such insight, corresponding roughly to an analytical or a practical inquiry into representation. An analytical inquiry attempts to provide a definition of representation that does justice to our intuitions and its central applications. A practical inquiry, by contrast, looks into the uses of the representations directly, and attempts to generalize, or at least figure out what a large number of its key applications may have in common. And, correspondingly, a theory or account of scientific representation may be substantial (if it aims for an analytical inquiry into its constituents) or deflationary (if its aim is instead to provide the most general possible account of its means in practice). In addition a substantial theory may be primitivist or reductive, depending on whether it postulates representation as an explanatory primitive, or attempts to analyze it away in terms of some more fundamental properties or relations. [4]

3.1. Substantial and Deflationary Accounts

On a substantial account every representation is understood to be some explanatory property of sources and targets, or their relation. Since on this analysis this property or relation is constitutive of representation, it is perdurably always there, as long as representation obtains. Its explanatory power suggests that it is the relation between the objects that make up possible sources and targets – or the type of properties shared, or the type of relation –, and not the singular properties of the concrete source-target pair, or their singular relations. The latter simply instantiate the type, and may vary greatly from application to application. By contrast, on a deflationary account there may be nothing but the singular properties of the concrete source-target pair in the particular representation at hand. On such an account nothing can explain why those singular properties (or relations) which are in fact used are the appropriate ones for representation – since there is, on a deflationary account, no constitutive explanatory relation of representation that they may instantiate. On the contrary on a typical deflationary account, more generally, the appropriate properties are exactly the ones that get used in the particular application at hand, and there is no reason why they are appropriate other than the fact that they do so get used.

Let me try to unravel this distinction a little further in analogy with theories of truth in metaphysics (as suggested in Suárez, 2004, p. 770; for an accessible review to theories of truth see Blackburn and Simmons, 1999). A substantial

[4] It is logically possible for a deflationary theory to also be reductive – i.e. to deny that representation is any substantial explanatory relation, but reduce every singular means of representation to some further non-substative properties. See section 6.2. for further discussion.

approach to truth assumes that there is a particular type of relation between propositions and facts (or between propositions and other propositions, or between propositions and utility functions) such that any proposition that stands in that relation to a fact (or to a set of other propositions, or some value of some agent's utility function) is true. This relation may be correspondence between propositions and facts; or coherence with other propositions; or the maximization of some utility function of an agent. On any of these substantial accounts, there is an explanation for why a proposition is true rather than false, namely that it so happens to hold such a relation (correspondence, coherence, utility) to something else. By contrast, on a deflationary account, nothing explains the truth of a proposition, since there is nothing substantial that it is true in 'virtue of' that could possibly explain it. To put it very bluntly, a proposition's truth is rather determined by its functional role in our linguistic practice. In other words, 'truth' picks out not a natural kind out there in the world, but a function in speech and discourse.

In a similar vein, substantial theories of representation understand it to be a substantial and explanatory relation akin to correspondence (or coherence or utility). Deflationary theories, by contrast, claim it to pick out nothing other than a functional role in scientific practice. Sections 5 and 6 review a few of each of those types of theories. A few distinctions need to be first introduced.

3.2. Reductive and Non-reductive accounts

The other important distinction already mentioned concerns reductive versus non-reductive theories of representation. A reductive theory is one that aims to reduce representation to something else – and most substantial accounts of representation will be reductive. Again one may think of the analogous case for theories of truth, where substantial accounts on the whole attempt to reduce the 'truth' property to a cluster of further properties that includes correspondence (or coherence, or utility). What explains the truth or falsehood of some proposition is then whether or not such properties obtain, since truth is just the obtaining of those properties. Similarly, a substantial account of representation will reduce it to some substantial and explanatory relation such as e.g. similarity. What explains representation is then the obtaining of this substantial property, since representation just is that property. On a reductive substantial theory, explanation comes for free.

A non-reductive account of representation, by contrast, will not attempt to reduce representation to anything else – it will not suppose that there exists any underlying property that explains away representation. Rather on this account representation is irreducible. Many deflationary accounts are non-reductive: they assume that representation cannot be analyzed away in terms of other properties. However, not all non-reductive accounts are deflationary, some are *primitivist*: they accept that the concept of representation cannot be reduced further, but they assume that this is so because it is an explanatory primitive – it can and usually is invoked in order to explain other concepts (for instance, it can be used to explain empirical adequacy, since this may be defined as accurate representation of the observable phenomena). In other words, a primitivist account accepts that representation cannot be reduced any further but not because it lacks substance. [5]

Hence, from a logical point of view, the substantial / deflationary and reductive / non-reductive distinctions are orthogonal. Any of the four combinations is logically possible. However, from a more practical point of view, substantial accounts of representation naturally line up with reduction, while deflationary accounts often, but not always, go along with the claim that representation cannot be further reduced or analyzed.

4. The Analogy with Art and Aesthetics

I have so far been elaborating on a novel analogy between theories of representation and theories of truth. There is a yet more conspicuous analogy running through the contemporary literature on representation linking it to discussions in aesthetics and the philosophy of art. [6] This analogy is often employed because similar views regarding artistic representation have been discussed for a long time (for an excellent treatment, see Lopes, 1996). Some consensus has been reached there regarding the strength of certain arguments against substantial theories of scientific representation and in particular resemblance theories. In this section I briefly review and expand on those arguments with an eye to an application later on in the context of scientific representation. The critical assumption must be that if representation is a substantial relation then it must be the same in both art and science (although the means of representation and the constraints on its application may vary greatly in both domains). Hence if representation is substantial it is so in both domains, while if it is deflationary then this must also be the case in both domains. This has plausibility because in both domains the representation in question is objectual, i.e. objects (models) stand for other objects (systems).

I review here only one important type of argument against resemblance theories of artistic representation, originally due to Nelson Goodman (1968), and which may be referred to as the *logical argument*. According to this argument, resemblance cannot constitute the relation of representation because it does not have the right logical properties for it. Resemblance is reflexive, symmetrical and transitive, but representation in general is none of this. This is best illustrated by a couple of paintings. Thus, Velázquez's "Portrait of Innocent X" depicts the Pope as he was sitting for Velázquez, but it certainly does not depict itself; and the Pope certainly does not represent the canvas. (Goodman uses the example of the portrait of the Duke of Wellington to illustrate the same 'Innocent' point). Francis

[5] It is unclear if there exists any primitivist account of truth, but Williamson's (2000) account of knowledge is certainly primitivist.
[6] The origins of the analogy are not always attributed correctly, or at all. It was originally introduced in Suárez (1999), which also contains some of the examples from art that have gone on to be discussed regularly. Van Fraassen (1994) is an acknowledged ancestor: although it does not present the analogy explicitly, it already discusses artistic representation in this context.

Bacon's "Study after Velázquez's Portrait of Pope Innocent X" (1953) is a formidable depiction of the Velázquez canvas, but it would be wrong to say that it depicts the Pope himself – thus exhibiting a failure of transitivity. Yet, to the extent that the Bacon resembles the Velázquez it also resembles the Pope; the Pope resembles the Velázquez just as much as is resembled by it; and, certainly, the canvas maximally resembles itself. In other words, resemblance is an equivalence relation: It is impossible to explain these failures of reflexivity, symmetry and transitivity if representation really is resemblance. Therefore resemblance [Res] cannot *constitute* representation in the sense of the definition: it is not the relation R such that "[for the canvas (s) and the Pope (t)]: R (s, t), and ii) for any source-target pair (S, T): S represents T if and only if R (S, T)". Here, only the first condition is fulfilled, but not the second. Now, certainly resemblance can *overlap* with representation – and indeed the Pope and the Velázquez canvas do (presumably!) resemble each other. Moreover, resemblance can be the effective means by which a viewer infers e.g. the colour of the Pope's clothes from the colour of the Velázquez canvas. So indeed the definition of means is satisfied at that very time for that particular viewer, since "[for the relation Res of resemblance, the canvas (s) and the Pope (t):] Res (s, t); and ii) some user of the representation employs Res (at that particular time and in that context) in order to draw inferences about t from s".

In other words, the analogy with artistic representation provides a logical argument against a substantial theory of the constituents of representation in general as resemblance (or indeed as any other relation that has the logical properties of reflexivity, symmetry and / or transitivity): It shows that representation in general is a not a logical equivalence relation.

5. Substantialism

Two main substantial accounts have been discussed in the literature, and they are both reductive in character. One of these accounts attempts to reduce representation to similarity relations between sources and targets while the other attempts a reduction to isomorphism between their structures. I argue below that both reductive attempts fail, and that it is an informative exercise to figure out exactly why. Nevertheless it bears reminding ourselves at this stage that one live option for the substantivist– one which moreover has not been sufficiently discussed in the literature so far – is to go 'primitivist' at this point, and deny that representation can in fact be reduced to any other property. I don't discuss this option here for the reasons of plausibility already mentioned, but it would certainly be a logically admissible way out of some of the arguments in this section.

5.1. Similarity

The connection between similarity and representation has been emphasized before (Aronson, Way and Harré, 1995; Giere, 1989; Godfrey-Smith, 2006; Weisberg, 2012) and in most of these cases it is at least plausible to suppose that the background assumption has been one of reduction. In other words all these accounts may be understood to be approximations to the following reductive theory of representation:

[sim]: A represents B if and only if A and B are similar

Any theory that has this form is a substantial account of the constituents of scientific representation that reduces it to the relation of similarity between sources and targets. In other words according to theories like this the underlying similarity between sources and targets explains their representational uses. Nevertheless the theories will differ in the different degrees of sophistication and complexity of their accounts of similarity. The simplest account understands similarity as the mere sharing of properties, and representation then boils down to the sharing of properties between representational sources and targets. Suppose then that the properties of the model source are given by $P^s = \{P^s_1, P^s_2, \ldots, P^s_n\}$, and the properties of the target object as $P^t = \{P^t_1, P^t_2, \ldots, P^t_m\}$. The simplest account then assumes that the source represents the target if and only if they share a subset of their properties. In other words, there are some $\{P^s_1, P^s_2, \ldots, P^s_i\} \in P^s$, with $i \leq n$, and some $\{P^t_1, P^t_2, \ldots, P^t_i\} \in P^t$, with $i \leq m$, such that $\{P^s_1 = P^t_1, P^s_2 = P^t_2, \ldots, P^s_i = P^t_i\}$. [7] On this account the complete objective positive analogy constitutes the representational relation between the billiard ball model and the gas molecules; while the negative analogy is a list of those properties of the model that play no genuine representational role. Hence only those properties of billiard balls that are shared with gas molecules, such as putatively elasticity and motion, are genuinely representational and can be said to be in the representational part of the model.

The simplicity of this account has a number of advantages, including its intuitiveness and fit with our ordinary or unreflective ways of talking about similarity. It moreover also fits in very well Mary Hesse's discussion of analogy, and her example of the kinetic theory of gases. But it has a number of problems too, which follow fairly straightforwardly from our earlier discussions. The most obvious problem is brought home by the analogy with art. The logical argument applies here in full force since similarity so simply construed is reflexive and symmetrical (and transitive over the properties shared by the intermediate targets). In other words, the simple account of similarity provides the wrong reduction property for representation.

The problem may be arguably confronted by more sophisticated definitions of similarity. I shall focus here on Michael Weisberg's (2012) recent account, which is heavily indebted to seminal work by Tversky and collaborators on the psychology of similarity judgements (Tversky, 1977). On the Tversky-Weisberg

[7] There is an issue, which I gloss over in the discussion in the main text, about how to define properties in such a way that these identity statements obtain. For instance, a purely extensional definition in principle does not help, since the entities that come under the extension of these properties are typically very different in the source and the target. I am therefore assuming that there is a different account of properties that makes sense of the identity statements above – otherwise this particular theory of representation as similarity would not even get off the ground.

account, a source *s* represents a target *t* if and only if their comparative similarity is large, where the degree of comparative similarity is measured in accordance to the following function: $Sim(s,t) = \theta\cdot f(S \cap T) - \alpha\cdot f(S - T) - \beta\cdot f(T - S)$, where S and T are the full set of salient and relevant features of source and target, and θ, α, β are relative weights which typically depend on context. This similarity measure is a function of shared features of source and target, which takes into account both features of the source missing in the target (i.e. Hesse's negative analogy) but also features of the target missing in the source. Moreover, and this is what is really nice about the Tversky-Weinsberg account, it renders similarity non-symmetric: *Sim (s, t)* need not equal *Sim (t, s)* because *s* and *t* may be endowed with a different range and number of salient features or attributes. So the Tversky-Weinsberg account provides an answer to the 'Innocent' point, and gets around the awkward conclusion that a model represents a target only to the extent that it is represented by it.

This more complex proposal is promising but it may be challenged on a number of grounds. Firstly, the Tversky-Weisberg ingenious measure of similarity – while going further than any similarity measure relying only on Hesse's positive and negative analogies – nonetheless cannot capture the impact of 'negative analogies by denial', i.e. the properties of the target that are explicitly denied in the source. We saw that Hesse does not distinguish those properties of the model source that fail to apply to the target (arguably colour and shine) from those other properties of the source that are explicitly denied in the target (limited elasticity, escape velocity). The former properties may perhaps be ignored altogether – since they do not play a role in the dynamical processes that ensue in either billiard balls or, naturally, gas molecules. The latter properties cannot however, be so dismissed – so it seems altogether wrong to claim that they are not part of the representation. In other words the similarity proposal does not account for a typical way in which models go wrong or misrepresent their targets. The form of misrepresentation that involves 'lying', 'simulating', or 'positively ascribing the wrong properties' is not describable under this account, which prima facie – given how pervasive 'simulation' is in model building – constitutes a problem for the account. [8]

Secondly, the logical argument has not been completely answered since Tversky-Weinsberg similarity continues to be reflexive – and so is therefore representation so construed, which seems just wrong. Secondly, notice the emphasis on contextual relevance. The sets of features of s, t to be taken into account are relative to judgements of relevance in some context of inquiry. In other words, there are no context-independent or context-transcendent descriptions of the features of sources and targets. Thirdly, and related to this, notice that the idea of context-relative description presupposes that some antecedent notion of representation is already in place, since it assumes that sources and targets are *represented as* having particular sets of features in context.

[8] One may suppose that this problem may be overcome by incorporating a fourth factor in the measure of similarity successfully representing the weight of features of the source that are explicitly denied in the target. It remains to be seen if this is possible, but at any rate the other three objections would remain.

5.2. Isomorphism

The other substantial theory appeals to the notion of isomorphism and its cognates. Once again there have been a number of different approaches (Bartels, 2006; Mundy, 1986; Pincock, 2012; Suppes, 2000; Swoyer, 1991), but they all seem to have at their foundation a commitment to reducing representation to some kind of morphism relation between the structures that are instantiated by both the source and the target. Therefore these views are all approximations to the following theory:

[iso]: A represents B if and only if the structures S_A and S_B exemplified by A and B are isomorphic: $S_A \equiv S_B$.

The definition of isomorphism is then given as follows. Two structures $S_A = \langle D_A, P^n_j \rangle$ and $S_B = \langle D_B, T^n_j \rangle$, where P^n_j and T^n_j are n-place relations, are isomorphic ($S_A \equiv S_B$) if and only if there is a one-to-one and onto mapping $f: D_A \rightarrow D_B$, such that for any n-tuple $(x_1,\ldots, x_n)$, where each $x_i \in D_A$: $P^n_j [x_1,\ldots, x_n]$ only if $T^n_j [f(x_1),\ldots, f(x_n)]$; and for any n-tuple $(y_1,\ldots, y_n)$, where each $y_i \in D_B$: $T^n_j [y_1,\ldots, y_n]$ only if $P^n_j [f^{-1}(y_1),\ldots, f^{-1}(y_n)]$. In other words, an isomorphism is a *relation preserving* mapping between the domains of two extensional structures, and its existence proves that the *relational framework* of the structures is the same. [9]

This theory is another substantial account of the constituent of representation; now as the relation of isomorphism between instantiated structures of the source and target pair. On this theory what explains representation is the conjunction of the obtaining of the relation of isomorphism and the appropriate relation of instantiation. [10]

Once again there are different versions of the theory appealing to different types of morphism relation, ranging from the strongest form (isomorphism) to weaker versions in this order: partial isomorphism, epimorphism, and homomorphism. I will not enter a detailed discussion of the differences, except to

---

[9] Or, as one also finds it sometimes said in the literature, it is an expression of 'structural identity'. But this is equivocal and should really be avoided, since the domains D_A, D_B are typically not identical but made up of very different elements. In other words, except in the reflexive case, isomorphism is certainly not an expression of numerical identity, as a cursory inspection of the definition makes obvious: "Two structures are isomorphic if and only if there is a mapping between their domains" can only be true if there exist two numerically distinct structures S_A and S_B endowed with their own distinct domains D_A and D_B.

[10] Thus what explains a particular item of representation is a product relation X.Y where X is isomorphism and Y is instantiation. There is an issue here with multiple instantiation, since X.Y = Z can be reproduced by an in principle infinite number of distinct such products, e.g. X'.Y' = X.Y= Z. Given that the extant leeway for instantiation turns out to be very large – most objects dramatically underdetermine their structure –, it follows that what precise isomorphism is constitutive of any particular representation is also almost always undefined.

state that the logical argument applies to some extent to all of them. For instance, isomorphism and partial isomorphism are reflexive, symmetrical and transitive; epimorphism and homomorphism are reflexive and transitive for the same set of relations, etc. More importantly still, for our purposes, is the fact that none of these approaches can accommodate the two senses of negative analogy. For there are no resources in any of these approaches to accommodate in particular the explicit denial of a relation in the target system that has been asserted in the source system. That is, whatever minimal morphism there is between two structures S_A and S_B, the relations that stand functionally related in the morphism cannot deny each other, i.e. $P^n_j [x_1,\ldots, x_n]$ and $T^n_j [f(x_1),\ldots, f(x_n)]$ cannot deny each other. At best in an epimorphism or homomorphism there will be some relations that are not asserted in either source or model: some $P^n_j [x_1,\ldots, x_n]$ will thus have no 'correlate'. But none of the known morphisms is able to account for a structural mapping where $P^n_j [x_1,\ldots, x_n]$ is explicitly denied in its correlate, i.e. where $T^n_j [f(x_1),\ldots, f(x_n)] = \neg P^n_j [x_1,\ldots, x_n]$. And this as we saw is precisely what happens in negative analogies "by denial", for instance when we model the perfectly elastic collisions of gas molecules by means of imperfectly elastic billiard balls. The only structural rendition of such analogies is one that does not insist on any transfer of relevant structure. Since most scientific modelling has this 'simulative' character so representation is not a structural relation even when model sources and targets are structures, or may be described as possessing or instantiating them (Pero and Suárez, Forthcoming).

6. Deflationism

Let us now look at deflationary theories. These are theories that do not presuppose that representation is substantial, or can otherwise be reduced to substantial properties or relations of sources and targets. The analogy with theories of truth was already noticed, and this also holds for deflationary views. Thus deflationary views of truth come in a couple of different forms, including what I call redundancy and use-based theories, and I argue below that deflationary theories of representation also take similar forms. In particular I develop a 'redundancy' version of RIG Hughes' Denotation-Demonstration-Interpretation (DDI) model of representation; and a 'use-based' version of my own inferential conception.

6.1. The 'DDI' account

Redundancy theories of truth originate in Frank Ramsey's work, in particular in "Facts and Propositions" (Ramsey, 1927). The basic thought is that to assert of some proposition 'P' that it is true is to assert nothing over and above 'P' itself. The predicate 'true' is instead redundant, in the sense that to predicate of any proposition that it is 'true' adds nothing to the content of that proposition. There is no substantial property that all true propositions share. The ascription of the predicate 'true' to a proposition is rather taken to possess only a kind of honorific value: it merely expresses the strength of someone's endorsement of a particular proposition. The use of the predicate may have other functions – for instance it helps in generalization and quotation – as in "everything Ed says is true" – but even then it does not express any substantial property – it does not establish that everything Ed says is true in virtue of any substantial property that all the propositions that he utters share. Truth is, if it is a property at all, a *redundant* property. For my purposes here I would like to focus on the part of the redundancy theory that most closely approaches the view that the terms 'truth' and 'falsity' do not admit a theoretical elucidation or analysis, but that, since they may be eliminated in principle – if not in practice – by disquotation, they do not in fact require such an analysis. I will take this implicitly to mean that there are no non-trivial necessary and sufficient conditions for these concepts.

The transposition of all this discussion to theories of scientific representation is, I argue, the claim that representation is itself a redundant concept in the same way: it expresses some honorific value that accrues to a particular use that agents make of some model, but it does not in itself capture any relational or otherwise property of sources or targets. The use of the term in addition signals further commitments, which I study in greater depth in the next section, and which are mainly related to the source's capacity to generate surrogate inferences regarding the target. But here again - as in the case of truth - such commitments do not signal that the term 'representation' picks out any substantial property. What they rather signal is that the term has no analysis in terms of non-trivial necessary and sufficient conditions. Concomitantly, we cannot work out explanations for the diverse applications of this term on the basis of any substantial definition, or precise application conditions, since it has none.

Perhaps the most outstanding example of a redundancy account of scientific representation is RIG Hughes' (1997) Denotation-Demonstration-Interpretation (or DDI) model.[11] On Hughes' account, representation typically although not necessarily involves three separate speech acts. Firstly, the denoting of some target by a model source; secondly, the demonstration internal to the model of some particular result; and thirdly, the interpretation of some aspects of the target as aspects of the source and, concomitantly, the transposition of the results of the demonstration back in terms of the target system with ensuing novel predictions, etc. The most important aspect of the DDI model for our purposes is Hughes' claim not to be "arguing that denotation, demonstration and interpretation constitute a set of speech-acts individually necessary and jointly sufficient for an act of theoretical representation to take place" (Hughes, 1997, p. 329). In other words, the DDI account is proposed explicitly as a deflationary account of representation along the lines of the redundancy theory: It refrains from defining the notion of representation, and it is not meant as an explanatory account of use. It is instead a description of three activities, or speech acts, that are typically enacted in scientific modelling. [12] It connects to practice in the sense that these items provide a description of three typical norms in the practice of representation.

---

[11] Van Fraassen (2008) too endorses a redundancy sort of deflationism regarding representation, but the details of his specific account are less developed.

[12] Hughes' original DDI model appealed essentially to the relation of denotation. I have argued that this appeal is problematic from a deflationary point of view, but that the original account may be amended into a fully deflationary one by replacing

The main example of the application of a DDI model is Galileo's model of motion along an inclined plane (Hughes, 1997, pp. 326-29), where a purely geometrical demonstration in geometry is taken to represent inertial motion on an inclined plane. Hughes shows convincingly how to apply the DDI 3-part speech act theory to this model and how much hinges in particular on the intermediate demonstration in the model. The DDI is of course supposed rather generally, so let us see to apply it to our main example, the billiard ball model in the kinetic theory of gases.

The denotation part of this model is prima facie straightforward: billiard balls are taken within the model to denote gas molecules. (There are some complications that arise when one considers the denotation of particular properties, particularly those in the negative analogy 'by absence', but we may leave property denotation aside for the purposes of the discussion). As for the demonstration and interpretation stages, Campbell (1920, pp. 126-128) helpfully separates what he calls the 'hypothesis' of the theory from the 'dictionary' provided by the model, which provides ideal grounds to apply both the demonstration and interpretation requirements in the DDI account. Thus among the hypothesis of the theory we find all the relevant mathematical axioms for the mechanical system, including those relating to the constants (l, m, v), the 3n-dependent variables (x_s, y_s, z_s) for the n-system of elastic balls, and the equations that govern them. The demonstrations will be carried out at this level via these mathematical equations. The dictionary brings in a link to the physics of gas molecules, by establishing e.g. i) that l is the length of the cubical vessel in which the 'perfect gas' is contained container, ii) m is the mass of each molecule, and nm the total mass of the gas, iii) $\frac{1}{\alpha}mv^2$ is the absolute temperature T of the gas, and iv) p_i is the pressure on the wall i of the container for i= x, y, z, which for the given interval of time comprised between t and t+γ is given by: $p_i = \lim_{\gamma\to\infty}\sum_{s=1}^{s=n}\frac{1}{\gamma}\Delta m\frac{d(x_s, y_s, z_s)}{dt}$. Then using the equations in the hypothesis of the theory we may calculate the pressure to be $p_i = \frac{1}{3l^3}nmv^2$, for any value of i. This constitutes the demonstration step referred to above. Now interpreting this result back into the physical description of the gas by means of the dictionary, we obtain: $p_i = \frac{\alpha\cdot n}{3}\frac{T}{V}$, which is an expression of Boyle's and Gay-Lussac's law since $\frac{\alpha\cdot n}{3}$ is a constant.

Hence the mechanical model generates, via the three steps of denotation (of gas molecules by infinitely elastic mechanical billiard balls), demonstration (by means of the hypothetical laws of the model) and interpretation (back in terms of thermodynamic properties of the gas) just the kind of prediction that is central to

---

denotation with something I call denotative function – a notion suggested by some of Catherine Elgin's writings, such as (2009) – which is not a relation but a feature of representational activity (Suárez, Forthcoming).

the kinetic theory of gases. The DDI account thus shows it to provide a *representation* of the gas and its properties.

6.2. The Inferential Conception

A different deflationary account is provided by the inferential conception (Suárez, 2004). It differs minimally from both RIG Hughes' DDI account and other deflationary views in leaning more towards use-based deflationism. [13] The difference between these two sorts of deflationism, however minimal, can again be illustrated by means of the analogy with theories of truth. All deflationary theories of some concept X deny that there is a definition of the concept that explains its use. Redundancy theories, as we saw, deny that X may be defined altogether; use-based theories admit that X may be defined, and may have possession conditions, but they deny that the use of X is thereby explained.

There is a sense then in which use-based views are reductive, although not substantial. They certainly aim to 'anchor' the concept in features of practice; and depending on how they go about this, they may in fact be reducing the concept. For instance, Giere (2004, p. 743) comes very close to providing necessary and sufficient conditions for sources to represent targets in terms of similarities, agents, and purposes – and may be considered reductive. The inferential conception that I have defended is more explicit in providing merely very general or necessary conditions. It does not aim at a reduction, but it does aim to link representation to salient features of representational practice – in a way that dissolves a number of philosophical conundrums regarding the notion.

On the inferential conception [inf] a source *s* represents a target *t* only if i) the representational force of *s* points to *t*, and ii) *s* allows an informed and competent agent to draw specific inferences regarding *t* (Suárez, 2004, p. 773). There are an important number of caveats and consequences to this definition, which may be summarized as follows:

a) [inf] may be objected to on the grounds of circularity, given the reference to representational force in part i). However, [inf] does not define representation and, at any rate, representational force is a feature of practice, not itself a 'concept'.

b) The term "representational force" is generic and covers all intended uses of a source *s* to represent a target *t*, including but not restricted to the notion of denotation whenever that may obtain. In particular while sources of fictional objects (such as the models of the ether at the end of the 19th century) do not denote them, they may well have representational force pointed at them.

[13] Ron Giere's (2004) recent 4-place pragmatic account is also arguably of this use-based deflationary kind.

c) Inferences are 'specific' if they lead to features of *t* that do not follow from the mere fact that *s* represents *t*. So, for instance, that someone's name is Sam follows from the fact that 'Sam' is used to denote him, so it is not specific – and the inference is therefore unable to ground representation.

d) A object or model is a representational source for some target within some representational practice as long as the practice in question involves norms of correct surrogate inference – it is not required in addition that any of the consequences of these inferences be to true conclusions. This allows the inferential conception to account for all instances of idealization and misrepresentation, including negative analogies 'by denial'.

e) The dynamics for developing scientific representations, on the inferential account, responds to the productive interplay of i) representational force and ii) inferential capacities.

The last point is particularly important to the prospects of the inferential conception, since any adequate account of scientific representation must do justice to this dynamical aspect. It arguably distinguishes the inferential conception from its more static substantive competitors (Suárez, 2004, p. 773-774). [14] It seems apposite to end this article with an outline of the application of the inferential conception to the same example of the billiard ball model in the kinetic theory of gases that has occupied us before. The main extension of the theory (discussed by both Campbell (1920, pp. 134-135) and Jeans (1940, pp. 170-174)) concerns viscosity in a gas. It does so happen that layers of the gas slide past each other as would be expected in a viscous liquid. This affects the relationship between molecular velocities and temperature, to the point that the coefficient α inserted into the hypothesis and related to temperature in the dictionary (see equation iii) above relating absolute temperature to molecular velocity) must be made to depend on the value of the temperature itself. In fact in a thorough treatment of the phenomenon, the assumption that the gas molecules are perfect elastic spheres must be relaxed (see Jeans (1940, p. 171) to deal with the fact that viscosity does not depend in actual fact – as may be established experimentally – on the size and shape of the container, but only on the temperature and density of the gas.

In other words the inferential capacities of the original version of the model lead, via its representational force, to predictions that in turn motivate an alteration in the model. This new model's reinforced inferential capacities avoid the potential refutation by adjusting the representational force of some of the elements in the model, notably viscosity, and so on. The conceptual development of the kinetic theory of gases is therefore accounted for in terms of the playing out of an inbuilt tension amongst the two essential surface features of scientific representation. The point certainly calls for further development and generalization, but the thought is that a deflationary account of scientific

[14] It moreover distinguishes it from other accounts that rhetorically claim to be in the same inferential spirit, yet ultimate presuppose that all inference rides upon some substantial structural relation. Contrary to appearances those accounts are, unlike the inferential conception, neither deflationary nor dynamical.

representation possesses the resources to naturally account for the heuristics that drive the dynamics of modelling practice.

7. Conclusions

Scientific representation continues to be a very lively area of research within contemporary philosophy of science. This article is intended to provide a review of some of the most significant work carried out in the area nowadays, while offering some critical commentary, guide, and direction. The first section introduced the topic of scientific representation from a historical perspective. The second section reviewed common terminology and drew some significant conceptual distinctions. In the third section accounts of representation were divided along two orthogonal dimensions into reductive and non-reductive, and substantial and deflationary. I argued that while there is no logical compulsion, it stands to reason that reductive accounts will be typically substantial while non-reductive ones will tend towards deflationism. Section four developed an analogy with artistic representation. In section five the major substantial accounts of representation were reviewed and it was argued that they all confront important challenges. Section six discussed and endorsed two delationary approaches, the DDI model and the inferential conception. Some promising avenues for future work in both deflationary approaches were finally suggested.

ACKNOWLEDGMENTS

Support from the Spanish Government (DGICT grant FFI2011-29834-C03-01) and the European Commission (FP7-PEOPLE-2012-IEF: project number 329430) is gratefully acknowledged.

REFERENCES

Aronson, J. E. Way and R. Harré (1995), *Realism Rescued*, London: Duckworth.

Bailer-Jones, D. (2006), *Scientific Models in Philosophy of Science*, Pittsburgh: Pittsburgh University Press.

Balzer, W., C. U. Moulines and J. Sneed (1989), *An Architectonic for Science: The Structuralist Program*, Dordrecht: Springer.

Bartels, A. (2006), "Defending the Structural Concept of Representation", *Theoria*, 55, pp. 7-19.

Black, M. (1962), *Models and Metaphors*, Ithaca, NY: Cornell University Press.

Blackburn, S. and K. Simmons, Eds. (1999), *Truth*, Oxford: Oxford University Press.

Boltzmann, L. (1902), "Models", Encyclopedia Britannica.

Brush, S. G. (2003), *The Kinetic Theory of Gases*, Imperial College Press.

Campbell, N. R. (1920), *Physics: The Elements*, Cambridge: Cambridge University Press.

Cartwright, N., T. Shomar and M. Suárez, “The Toolbox of Science”, *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 44, pp. 137-149.

Duhem, P. (1906/1954), *The Aim and Structure of Physical Theory*, Princeton: Princeton University Press. (English translation of original 1906 book in French).

Elgin, C. (2009) “Exemplification, Idealization and Understanding”, in M. Suárez (Ed.), *Fictions in Science: Philosophical Essays on Modelling and Idealization*, Routledge, 2009, pp. 77-90.

Giere, R. (1988), *Explaining Science*, Chicago: University of Chicago Press.

Giere, R. (2004), “How Models are Used to Represent Reality”, *Philosophy of Science* 71, pp. 742-752.

Godfrey-Smith, P. (2006), “The Strategy of Model-based Science”, *Biology and Philosophy*, 21, 725-740.

Goodman, N. (1968), *Languages of Art: An Approach to a Theory of Symbols*, Indianapolis: The Bobbs-Merril Company.

Harré, R. (1960), *An Introduction to the Logic of the Sciences*, London: MacMillan Press.

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

Hughes, RIG (1989), *The Structure and Interpretation of Quantum Mechanics*, Cambridge, MA: Harvard University Press.

Hughes, RIG (1997), “Models and Representation”, *Philosophy of Science*, 64, pp. 325-36.

Hughes, RIG (1999), “The Ising Model, Computer Simulation and Universal Physics”, in Morgan and Morrison, Eds., *Models as Mediators*, pp. 97-145.

Jeans, J. (1940), *Kinetic Theory of Gases*, Cambridge: Cambridge University Press.

Morgan, M. and M. Morrison, Eds. (1999), *Models as Mediators*, Cambridge: Cambridge University Press.

Mundy, B. (1986), “On the General Theory of Meaningful Representation”, Synthese, 67, pp. 391-437.

Pero, F. and M. Suárez (Forthcoming), “Varieties of Misrepresentation and Homomorphism”.

Pincock, C. (2012), *Mathematics and Scientific Representation*, Oxford: Oxford University Press.

Toulmin, S. (1990), *The Philosophy of Science*, Harper Torchbooks.

Ramsey, F. P. (1927), “Symposium: Facts and Propositions”, *Proceedings of the Royal Society Supplementary Volumes*, vol. 7, pp. 153-206.

Suárez, M. (1999), “Theories, Models and Representations”, in L. Magnani et al. (Eds), *Model-Based Reasoning in Scientific Discovery*, Kluwer University Press, pp. 83-95.

Suárez, M. (2003), “Scientific Representation: Against Similarity and Isomorphism”, *International Studies in the Philosophy of Science*, 17, pp. 225-244.

Suárez, M. (2004), “An Inferential Conception of Scientific Representation”, *Philosophy of Science*, 71, pp. 767-779.

Suárez, M. (2014), “Scientific Representation”, Oxford Bibliographies Online.

Suarez, M. (Forthcoming), *Inference and Representation*, Chicago: University of Chicago Press.

Tversky, A. (1977), “Features of Similarity”, *Psychological Review*, 84, pp. 327-352

Suppes, P. (2000), “Representation and Invariance of Scientific Structures”, Stanford CSLI series, University of Chicago Press.

Swoyer, C. 1991). “Structural Representation and Surrogative Reasoning”, *Synthese* 87, pp. 449-508.

Van Fraassen, B. (1980), *The Scientific Image*, Oxford: Oxford University Press.

Van Fraassen, B. (1994), “Interpretation of Science, Science as Interpretation”, in J. Hilgevoord (Ed.), *Physics and Our View of The World*, Cambridge: Cambridge University Press, pp. 188-225.

Van Fraassen, B. (2008), *Scientific Representation*, Oxford: Oxford University Press.

Weisberg, M. (2012), *Simulation and Similarity*, Oxford: Oxford University Press.

Williamson, T. (2000), *Knowledge and its Limits*, Oxford University Press.