

Introduction

José DÍEZ and Roman FRIGG

BIBLID [0495-4548 (2006) 21: 55; p. 5]

Many philosophers, realists and antirealists alike, agree with a characterisation of science as an activity aimed at representing (selected aspects or parts of) the world. But what does it mean to scientifically represent something? The contributions to this special issue approach this question from different angles. They naturally fall into two groups: the first three papers defend particular accounts of scientific representation while the latter two take issue with influential positions.

In ‘Defending the Structural Concept of Representation’ Andreas Bartels argues that scientific representation is based on homomorphism and addresses different criticisms that have been levelled against this view, which leads him to introduce the distinction between potential and actual representations. Andoni Ibarra and Thomas Mormann’s ‘Scientific Theories as Intervening Representations’ takes up ideas going back to Pierre Duhem and Heinrich Hertz and develops a theory construing representations as complex commutative graphs, which serve as the basis for a discussion of the *in vivo/in vitro* problem in biochemistry. Mauricio Suárez and Albert Solé, in ‘On the Analogy Between Cognitive Representation and Truth’, point to communalities between the minimalist conception of truth and their own pluralist account of cognitive representation, from which they muster support for a deflationary attitude towards scientific representation in general.

In ‘Scientific Representation and the Semantic View of Theories’ Roman Frigg first introduces three problems that every account of scientific representation has to come to terms with and then argues that the widely-held model-theoretic approach to theories does not provide a valid response to any of them. The last paper of this special issue, Craig Callender and Jonathan Cohen’s ‘There Is No Special Problem About Scientific Representation’ offers a radically sceptical perspective on the entire debate by arguing that scientific representation is only a special case of a more general notion of representation, and that nothing over and above a well worked-out theory of the latter is needed to account for what happens in the sciences.

Finally, we would like to thank the editor of *Theoria* for supporting this project and Andrew Goldfinch for his invaluable assistance in the production of the final version of the papers.



Defending the Structural Concept of Representation

Andreas BARTELS

BIBLID [0495-4548 (2006) 21: 55; pp. 7-19]

ABSTRACT: The aim of this paper is to defend the structural concept of representation, as defined by homomorphisms, against its main objections, namely: logical objections, the objection from misrepresentation, the objection from failing necessity, and the copy theory objection. The logical objections can be met by reserving the relation 'to be homomorphic to' for the explication of *potential* representation (or, of the representational content). *Actual* reference objects ('targets') of representations are determined by (intentional or causal) representational mechanisms. Appealing to the independence of the dimensions of 'content' and 'target' also helps to see how the structural concept can cope with misrepresentation. Finally, I argue that homomorphic representations are not necessarily 'copies' of their representanda, and thus can convey scientific insight.

Key words: (structural concept of) representation, homomorphism, content.

1. Introduction

What is the essence of representation? This question has motivated a lively debate within both philosophy of science and the cognitive sciences. Discussions of this issue center on two questions. First, is the concept of representation appropriate and useful for the study of cognitive processes? Second, can representation, in general, be understood as a transfer of structure, from some original domain to some representing domain? The focus of this paper is on the second question. The structural concept of representation, as I call it in this paper, has been advocated by Mundy (1986), Watson (1995), Swoyer (1991) and French (2003), but resolutely rejected by Goodman (1976), Scholz (1991), Grush (1995), Hughes (1997), Suárez (2003, 2004), Bailer-Jones (2003) and others. Generally, those who reject the structural concept of representation do so for conceptual, not empirical, reasons (an exception is Grush (1995)). In what follows, I shall defend the structural concept of representation by demonstrating that the conceptual objections can be refuted. After introducing the structural concept of representation (section 2), I shall challenge the main objections against the structural concept of representation: logical objections (section 3), the objection from misrepresentation (section 4), the objection from failing necessity (section 5), and the copy theory objection (section 6).

2. The Structural Concept of Representation

The structural concept of representation claims that something, B , can *represent* something, A , only if some structure of the represented domain A is transferred to its image B . To make this idea more precise, A (the domain to be represented) and B (the



domain representing \mathcal{A}) are described by *similar¹ relational structures*. A relational structure is given by a set, on which one- up to n -place relations $R_1^{\mathcal{A}} \dots R_m^{\mathcal{A}}$ (respectively $R_1^{\mathcal{B}} \dots R_m^{\mathcal{B}}$) are defined. Now a mapping $f: \mathcal{A} \rightarrow \mathcal{B}$ can be defined which maps \mathcal{A} onto \mathcal{B} . The mapping f is not necessarily one-to-one and satisfies two following conditions:

- (i) For all j and all elements a_i of \mathcal{A} : if $R_j^{\mathcal{B}}(f(a_1), \dots, f(a_n))$, then $R_j^{\mathcal{A}}(a_1, \dots, a_n)$

Condition (i) requires that for all relations $R_j^{\mathcal{B}}$, if some images $f(a_1) \dots f(a_n)$ of the arguments a_1, \dots, a_n under f satisfy the relation, then the arguments also satisfy the corresponding relation $R_j^{\mathcal{A}}$ on \mathcal{A} . If that is the case, then f is called a *faithful* mapping of \mathcal{A} onto \mathcal{B} . Representations should be modeled by means of faithful mappings. Otherwise there may be facts in the representing domain to which there are no corresponding facts in the represented domain.

The second condition is that the facts in \mathcal{B} give *complete* information about facts in \mathcal{A} , that is, for every fact in \mathcal{A} there must be a corresponding (representing) fact in \mathcal{B} .

- (ii) For all j and all elements a_i of \mathcal{A} : if $R_j^{\mathcal{A}}(a_1, \dots, a_n)$, then $R_j^{\mathcal{B}}(f(a_1), \dots, f(a_n))$

If (i) and (ii) are fulfilled, f is a homomorphism from \mathcal{A} onto \mathcal{B} , and \mathcal{B} , by virtue of the existence of f , can be said to be an *homomorphic image* of \mathcal{A} (Dunn and Hardegree 2001, 15). The structural concept of representation claims that \mathcal{B} represents \mathcal{A} only if \mathcal{B} is a homomorphic image of \mathcal{A} . (In the following, if \mathcal{B} is an homomorphic image of \mathcal{A} , I will say that ‘ \mathcal{A} is homomorphic to \mathcal{B} ’).

In section 3, I shall discuss a differentiation concerning ‘ \mathcal{B} represents \mathcal{A} ’ that turns out to be a necessary reaction to the logical objections against the structural concept of representation. I shall then introduce two independent components of the relation of representation: the representational *content* and the *target* of the representation. Accordingly, ‘ \mathcal{B} represents \mathcal{A} ’ can either mean ‘ \mathcal{A} is a part of the representational *content* of \mathcal{B} ’ or ‘ \mathcal{A} is the target (reference object) of \mathcal{B} ’. Understood in the first sense, ‘ \mathcal{B} represents \mathcal{A} ’ is to be explained by the relation of \mathcal{A} being homomorphic to \mathcal{B} . However, the component of the relation of representation in the second sense cannot be understood by means of homomorphisms; I shall explain this in more detail in section 3.

In section 6, I shall discuss the reasons for employing homomorphism rather than isomorphism for modelling the representation relation. (Indeed, the failure to distinguish between homomorphism and isomorphism may be the main source of the copy-theory misunderstanding with respect to homomorphisms). For now, I shall merely cite the example given by Dunn and Hardegree (2001) for the case of homomorphism. They note that a photographic image (a clear case of representation) ‘is not isomorphic to its subject, even in the ideal, for at least the following reasons: (1) the image is two-dimensional, whereas the subject is three-dimensional; (2) the image depicts

¹ Two relational structures \mathcal{A} and \mathcal{B} are similar, if they are of the same type, that is, if all corresponding relations on \mathcal{A} and \mathcal{B} have the same number of arguments (see Dunn and Hardegree 2001, 10).

only the surface of the subject, whereas the subject presumably has inner detail, not conveyed by the image; (3) the image may be in black and white, whereas the subject is presumably ‘in color’ (Dunn and Hardegee 2001, 15). As the example shows, to take isomorphisms as the core of representation would fail even with regard to some very common instances of representation.

On the other hand, it has to be admitted that ‘homomorphism’ is a very general notion, which has to be filled out by some specific types of mappings to model concrete cases of representation. The objects of homomorphic representations can, for instance, be perceptual objects. A specific type of relations, defined on the representing domain, are geometrical structures (distance structures on vector spaces) as they are used by Peter Gärdenfors to model conceptual representations (see Gärdenfors 2000), or mereological structures, as they may be useful to model non-conceptual representations. In all these cases, different concrete types of relations specify the relations defining the relational structures, but the claim in all the cases is that structures are transferred from the represented to the representing domain.

Homomorphisms, as defined above, describe an idealized case. The conditions that hold for homomorphisms can be weakened so as to fit the cases in which representations do not work perfectly. This can happen with respect to two criteria: faithfulness and completeness.

The faithfulness that is required in the definition of homomorphisms is in a sense ‘absolute’, as the satisfaction of the corresponding relation in the represented domain A is required with respect to *all* ur-images of $f(a_1), \dots, f(a_n)$. (Since f is not necessarily one-to-one, there can be more than one ur-image.). The absolute condition of faithfulness for the information the representation provides about the represented domain may be weakened to the restricted notion of ‘*minimal fidelity*’. This weaker notion only requires that *there are* ur-images which satisfy the corresponding relation defined on A .² Compared to absolute faithfulness, this notion ensures that for every fact in B there is a corresponding fact in A . If a representational mechanism fulfils only minimal fidelity, the representation may lead to false expectations concerning facts in the represented domain A . For instance, the visual system of an organism indicates directions of stimuli in the visual field only up to a range of fuzzyness; the representational mechanism then produces *non-exact* representations (as it is to be expected for representational mechanisms in the biological world).

Non-exact representations blur some of the fine grained differences existing in the represented domain. In other words, their representational content does not reflect those differences. In order to be able to describe the representational content even in those cases by means of the transduced homomorphic structure, we have to ‘adapt’ the represented domain A by identifying all arguments in A which are mapped to the same element of B by the function f ; thereby the old arguments of f are replaced by new arguments which are equivalence classes of old arguments (the equivalence relation being the relation of ‘being mapped to the same element of B ’). By this identifica-

² This condition is called ‘minimal fidelity’ by Dunn and Hardegee 2001, 17.

tion procedure the description of representational content by means of transduced homomorphic structure can be restored. In the extreme, the representation blurs all differences existing in A . Then the representation has degenerated into a detector of A -like events.

The conditions that hold for homomorphisms can also be weakened with regard to *completeness*. In the ideal case, facts about A are preserved by facts about B ‘for all j ’, i.e. for all relations that are defined on A . In less ideal cases, there may only exist *some relations* for which this preservation holds. In one type of weakening the homomorphism can be trivially eliminated by cutting off from the relational structure A all relations for which the homomorphism-condition is not satisfied.³ A second type of weakening allows a representation to represent a certain property or relation only for a *limited range of arguments*.⁴ For example, the representational capacity can be limited to a certain region of stimuli from the environment. The visual system of an organism, for instance, is sensitive only for a certain range of wave lengths.

The fewer relations for which the transfer of structure holds, and the fewer the number of elements of A to which the transfer is restricted, the poorer the representation will be with respect to content. In an extreme case, no content will be left.

3. Logical Objections

Many critics of the structural concept of representation think that the concept, first of all, fails to meet the most obvious logical conditions for defining representation. One difficulty evaluating those objections is that the critics’ attacks are directed against what they call the ‘similarity theory of representation’ (Goodman 1976) or the ‘isomorphism conception of representation’ (cf. Scholz 1991,⁵ Suárez 2003, 2004). Simi-

³ In the same vein, Swoyer (1991, 470f.) discusses cases of representations in which the mapping between the represented and the representing domain ‘does not respect all of the relations in the original system, but only some’. One of his examples is the two-dimensional projection of a sphere that ‘cannot depict all of its features without distortion, so when we use flat maps to represent the Earth, something has to give’. The structural theory of representation can be accommodated to those cases, according to Swoyer, by restricting the operation of the representing function to the respected relations (Swoyer 1991, 472). French and da Costa (2003) have coined the notion of a ‘partial isomorphism’ to describe representation relations, which are restricted in their scope to a certain substructure of a given structured domain. For instance, the representation provided by the billiard ball model for gases, as described by the kinetic theory of gases, is a representation “in certain respects and to certain degrees” (French and da Costa 2003, 49), and therefore has to be described by partial isomorphisms operating on partial structures.

⁴ This sort of weakening of the homomorphism conditions is discussed by Swoyer (1991, 470/71): ‘In some cases of representation, relations are respected only under certain conditions (e.g., boundary conditions). For example, a mercury thermometer may reliably represent the temperature if it is neither too hot nor too cold, but it would fare poorly in liquid helium or near the surface of the sun.’

⁵ According to Scholz, isomorphisms not only lack adequate logical properties for explaining the representation relation, but also is in danger of trivialization (see Scholz 1991, 59). According to this argument, it is possible to define an isomorphism between arbitrary complexes. Now, if my television set, by means of its being isomorphic to my body, were a representation of my body, the whole structural

larity, or isomorphism, it is argued, lack adequate logical properties for explaining representation relations, because (i) representation relations are *non-reflexive*, whereas similarity (and isomorphism) are reflexive, and (ii) representation relations are *non-symmetric*, whereas similarity (and isomorphism) are symmetric. Since ‘to be homomorphic to’ is non-symmetric, only objection (i) can also be raised automatically against the homomorphism theory of representation. Nevertheless, objection (i) suffices to show that in general, for A to be represented by B , it would not be *sufficient* that A is homomorphic to B . There has to be some additional component of the representation relation, which prevents that each object, endowed with a relational structure, represents *itself*.

The *locus classicus* for logical objections against the structural concept of representation is Goodman’s *Languages of Art*.

The most naive view of representation might perhaps be put somewhat like this: “A represents B if and only if A appreciably resembles B”, or “A represents B to the extent that A resembles B”. Vestiges of this view, with assorted refinements, persist in most writing on representation. Yet more error could hardly be compressed into so short a formula. Some of the faults are obvious enough. An object resembles itself to the maximum degree but rarely represents itself; resemblance, unlike representation, is reflexive. Again, unlike representation, resemblance is symmetric: B is as much like A as A is like B, but while a painting may represent the Duke of Wellington, the Duke doesn’t represent the painting. Furthermore, in many cases neither one of a pair of very like objects represents the other: none of the automobiles off an assembly line is a picture of any of the rest; and a man is not normally a representation of another man, even his twin brother. Plainly, resemblance in any degree is no sufficient condition for representation. (Goodman 1976, 3-4)

Goodman makes three points: the reflexivity and symmetry objections, and he notes that objects resembling each other do not necessarily also *represent* each other. Goodman’s criticism is, with the exception of the symmetry objection (ii), also appropriate with respect to the homomorphism theory of representation. The proponent of the homomorphism theory has to admit that the extension of ‘to represent’ is at most a *proper subset* of the extension of the relation ‘to be homomorphic to’, and he is obliged to explain *why* the extensions do not coincide.

In order to explain why the extensions do not coincide, I shall introduce the distinction between *potential* representations and *actual* representations. B is a potential representation of A , if B can be used to correctly represent A , given the existence of some representational *mechanism* connecting A with B . I will say, then, that A is part of the representational content of B . For example, one can use a road map to correctly

similarity theory would be trivialized. The force of this argument is not very strong. Of course, it will be possible to identify the same number of parts in both of the complexes, the television set and my body, and it will also be possible to impose relations on these complexes such that there is a one-to-one mapping between them, preserving the relations. For instance, simply take the relation of ‘being identical with’ for both complexes. Now the claim is that my television set would have to be accepted as a representation of my body with respect to this relation. I have no problem with admitting this, since this merely means that there exist rather uninteresting forms of representation. What makes a representation interesting is that the preserved relations tell us something important about the object to be represented with respect to the representational context.

represent one's way home, if one intentionally takes a certain red curve on the map to stand for the highway which one has to pass etc. Since the road map is endowed with the relevant structure, it entails a potential representation of his or her way home that can be exploited by means of an intentional representational mechanism. Thus, I shall claim that A being homomorphic to B is sufficient for A to be potentially represented by B , i.e. that the extensions of the relations 'to be homomorphic to' and 'to represent potentially' coincide. In order for B to be also a correct *actual* representation of A , A has to be selected as the *target* of the representation from the set of objects potentially represented by B (i.e., from the content of B) by some representational mechanism connecting A with B . If there is a representational mechanism connecting A with B , but B is not a potential representation of A , then B *misrepresents* A .

The most important sorts of representational mechanisms are *representational intentions* and *causal relations*. The existence (or non-existence) of representational mechanisms of these sorts explains how we get (or fail to get) actual representations out of potential ones. For example, a chair, considered as a structured object formed by its parts, is homomorphic to itself, and therefore potentially represents itself. But in this case a representational intention has to occur, in order to turn this potential representation into an actual representation. The chair usually does not represent anything, but serves a certain purpose. On the other hand, if you find the chair being an exhibit in an art exhibition, then the idea is not too far fetched that the chair might be an (actual) representation of something. Now if you ask what kind of thing the chair represents, it may turn out that the artist intends, by placing the chair in a certain place or way, that the chair (reflexively) represents *itself*, in order to suggest to the visitor of the exhibition that a thing is not necessarily what it is most of the time, namely something made for a certain purpose, but that it can also be seen as 'standing for itself'. The intention of the artist figures as a representational mechanism turning a potential representation into an actual representation. Thus, Goodman was right to insist that, as a matter of fact, 'an object rarely represents itself'. But we now see that this does not count against a structural concept of representation. Instead, it brings to our attention the requirement that to make something an *actual* representation of itself representational mechanisms are needed, and that this requirement is rarely satisfied.

Causal representational mechanisms are exemplified by photography. A photograph may be a potential representation of my son, and *vice versa*. But it is an actual representation only in one direction, because there is a one-way causal process connecting the light rays emanating from my son's body and resulting in the photography, but not the other way round. Nevertheless, there may be 'irregular' contexts, in which representational intentions do not follow the regular causal direction. Again, Goodman is right in insisting 'the Duke doesn't represent the painting.' But in some odd contexts, the Duke could nevertheless be seen as representing the painting of the Duke. Imagine, for example, an impoverished Duke who is now forced to imitate the painting on fairs. Perhaps the painting has become famous during the Duke's personal decline. Whereas in the common context of portrait painting only pairs of the relation are used to instantiate actual representations in which the first element is a person, and

the second a picture, a pair with the reversed order fulfils the actual representation relation in this strange example.

Causal or intentional representational mechanisms also determine *how the relational structures* that are related by the homomorphism are defined: this includes the identification of the elements of the base set and the relations that are seen as relevant for being represented with respect to the context in question. I refer to all this as the *pragmatic* conditions of actual representations (see Bailer-Jones 2003). My concern, in this paper however, is potential representation as the necessary condition of correct actual representation. It was my aim in this section to show that the standard logical objections against the structural concept of representation don't have the force to exclude a homomorphism conception of potential representation. The structural properties of an object determine what the object potentially represents. If these representational resources are exploited by intentional users or by causal processes, then actual representations emerge.

4. The Objection from Misrepresentation

A second general objection against the homomorphism theory of representation is that homomorphisms do not allow for misrepresentations. Misrepresentation is a common empirical phenomenon, thus no concept of representation will be empirically adequate without being able to explain it. What is more, permitting misrepresentation is a condition of *conceptual adequacy*, since the very concept of representation presupposes the possibility of a distinction between the case in which some X *misrepresents* some Y and the case, in which X does *not* represent Y at all.

Why are homomorphisms perceived to be unable to fulfill this condition? The reason is that a homomorphism between relational structures A and B either exists or does not exist; in the first case, B represents A , whereas in the second case B does *not* represent A . What would it mean for B to represent A , but *incorrectly*?

Contrary to first appearances, the homomorphism theory does not have problems to allow for misrepresentation. If B represents A , then B *refers* to A . There is also a *content* of that representation which is not necessarily identical with its reference. B misrepresents A just in case B refers to A but the representational content does not entail A . Intuitively this means that B is about A , but does not match A in what it says about A . Problems with misrepresentation arise because some theories of representation do not have the resources to identify reference and content independently. If and only if reference and content are conceptually identified do we lack resources to explain misrepresentation. The following demonstrates that this is not the case for the homomorphism theory:

- (a) According to the homomorphism theory the reference of B is A iff A is the target of B , which is determined by a representational mechanism.⁶

⁶ The notion of a 'target' follows Cummins (1996). But whereas in Cummins' sense, the target is determined by the intended use a certain representational state is supposed to serve, I assume that the ref-

(b) The content of B entails A iff A is homomorphic to B .

Hence reference and content do not coincide from which it follows that misrepresentation is possible

This shows that the impression of the theories' inability to cope with misrepresentation arises from the wrong assumption that by the existence of a homomorphism between A and B the *reference* of B would be determined. This assumption cannot be right, simply because no unique reference object for B can be determined on the basis of the property of being homomorphic to B . Therefore, the determination of a reference object for B has to be explained independently of the homomorphism theory. The reference of B is fixed by a representational mechanism, i.e. either by an intentional or a causal process (cf. section 3). In contrast, the representational *content* of B is determined by B 's structural properties, i.e. by the relational structure B is endowed with. This relational structure of B determines what objects (which are themselves relational structures) are homomorphic to B , i.e. it determines the set of objects that are *potentially represented* by B (cf. section 3). Thus, the representational content of B is identical with the set of objects that are potentially represented by B . Reference objects for B will be determined independently of the content of B . If a reference object for B is chosen by a representational mechanism out of the set of objects potentially represented by B , then B will correctly represent this object. If a reference object for B is chosen which does not belong to this set, then this reference object will be *misrepresented* by B . Thus, the case in which something A is misrepresented by B and the case in which A is *not* represented by B (i.e. A is not a reference object of B) are clearly distinct. This means that the homomorphism theory of representation has the resources to explain misrepresentation.

A nice example of the ability of the homomorphism theory to explain misrepresentation is Cummins' (1996) example of a chess computer. The calculations of the chess computer are intended (by the constructor) to result in a certain position, in response to the moves of the computer's opponent (this position is the *target*, or the *reference object*), but —by some failure in the computer's architecture— the computer performs by indicating a different position. The relevant structure of this different position is the *representational content* and the appearance of a difference between the two positions means that a misrepresentation has occurred (see Cummins 1996, 5f.).

5. *The Objection that Homomorphism is not Necessary*

Even if it is accepted that the structural concept is able to overcome both the logical and the misrepresentation objections, this may only mean that *some* representations may exhibit homomorphic structure. But perhaps homomorphisms are not necessary for representations? Indeed, most critics (e.g. Scholz 1991, Suárez 2004) argue that homomorphism is neither sufficient nor necessary for representation. Instead of going

erence could also be fixed causally. I do not intend in this paper to give any reasons that could favour one theory of reference against some other theory.

into the details of their argumentation, with regard to this objection, I prefer to consider two types of phenomena that are, as far as I can see, the most obvious paradigms of representation that allegedly work without homomorphism. In the first case, the result will be that representations are involved, but (contrary to first appearances) invoke homomorphisms, whereas in the second case, it will turn out that no representations are involved. The first paradigm type is detectors that have no internal structure.

There are both natural and artificial systems using some detector systems in order to represent certain conditions occurring in their environment, but the detectors cannot be interpreted as being homomorphically related to the represented condition. Examples are the detection of the direction of the magnetic field by the magnetosomes of sea bacteria, or neuronal systems in the human retina that are able to detect certain directions of moved stimuli occurring in the visual field (see Goldstein 1996, 274). The representation,⁷ in the latter example, is performed by ‘yes’- or ‘no’-answers which are produced by the neuronal system corresponding to whether the stimulus occurs within a given range of spatial directions or not.

The answer given by the detector system depends on whether the direction of the stimulus occurring in the visual field is such as to generate neuronal inputs which are transduced through the system up to the central neuron. The firing of the central neuron then means a ‘yes’-answer of the system with regard to the stimulus. In that case, the system represents the corresponding type of stimuli. If the inputs inhibit each other, the signal is cancelled out with the result that no firing of the central neuron occurs. This means a ‘no’-answer of the system. It is crucial for the answer of the system, in which direction the stimulus moves through the visual field. The direction determines, in what temporal succession the neuronal inputs enter the detector system, and the succession determines whether the inhibiting neuronal connections are able to cancel out the signal or not. For each direction, there is an optimal detector, such that to stimuli in that direction it is maximally responsive, whereas stimuli moving in the opposite direction are completely ignored by the detector.

Whether a certain type of stimuli is represented by the neuronal system, depends on whether its direction fits the *internal structure* of the system. The detector system is like a lock in that a key (the direction in which the stimulus moves) fits or not. Since the internal neuronal structure of the detector explains its representational performance, the internal structure of some detector may be seen as determining its representational *content*. But, there is no internal structure of some single stimulus. Thus, a ‘yes’-answer of the system corresponding to that stimulus cannot be explained by the existence of a homomorphism between the stimulus (described as a relational structure) and the neuronal system. This means that a single neuronal detector *does not represent* a single stimulus by means of the stimulus being *homomorphic* to the detector. The

⁷ Neuronal detectors of direction stimuli are typical examples of representations described in studies of human and animal visual systems. I will presuppose that such examples have to be understood as representations by any serious theory of representation.

single neuronal detector has no representational content in the sense of the homomorphism theory.

Representation occurs in this case in the first instance not between the single concrete entities but between the stimulus space S and the detector-space D , by means of a homomorphism d operating between the space of all types of stimuli (each defined by a certain direction a) and the corresponding detectors $d(a)$. The stimulus-types are related by rotations with respect to the horizontal plane through the retina, and the same applies to the corresponding detector systems. Thus, the homomorphism d preserves the connection structure of mappings (rotations) o on S , since every mapping on S corresponds to a mapping o^* on D (which is a rotation of detectors):⁸

$$\text{For rotations } r_1 \text{ and } r_2 \text{ on } S: d(r_1 \circ r_2) = d(r_1) \circ^* d(r_2).$$

The representation that holds in the first instance between the spaces induces a representation in a derivative sense also between single stimuli and single detectors by means of their occupying corresponding places in the homomorphic mapping structures.

The second paradigm case of alleged representations that work without homomorphisms are intentional denotations by arbitrary signs, be they pencils representing kings in a child's play, coins standing for soccer players used to demonstrate a possible move, or what ever. Arbitrary signs *denoting* objects, phenomena, situations, or types of behavior, are often held to be *the* paradigmatic instances of representation. Indeed, an arbitrary sign has something of a representation, insofar as some intentional representational mechanism (a decision about its denotation) has assigned it to a reference object. But an arbitrary sign cannot *misrepresent*. The reference object to which the sign has been assigned by an intentional act cannot *fail* to belong to the set of objects potentially represented by the sign, that is, it cannot fail to belong to the representational content of the sign. The reason is that the sign simply *has* no representational content. In order to have representational content conceptually independent of its reference object, there would need to be properties of the sign delimiting the set of objects the sign could be correctly used to represent. Since, by definition, no property of an arbitrary sign has any representational relevance (beside the denotation act applied to it), arbitrary signs cannot misrepresent. Since, for any conception of representation, permitting misrepresentation is a condition of *conceptual adequacy* (cf. section 4), arbitrary signs, contrary to first appearances, are not representations.

Even Nelson Goodman, supposedly one of the main advocates of the denotation view, has actually been very sceptical about denotation as a means of representation.

⁸ Representations that work by transducing the connection structure of the represented domain to the representing domain are discussed by Ibarra and Mormann (2000) under the notion of *homology*. Contrary to what is claimed by Suárez (2004, 769), the homology theory explains representation as structural relation between the represented and the representing domain; both, the represented and the representing domain are conceived of as algebraic structures, namely as group structures of morphisms, and the relation between them is 'structural', since it is a homomorphism operating on the group structures.

True, Goodman argues that '[d]enotation is the core of representation and is independent of resemblance' (Goodman 1976, 5). This means that the *reference* of a representation cannot be determined by similarity, but has to be conceived as a primitive relation. On the other hand, he notes that representations, for instance paintings, do have a *content* that is independent of the denotation (reference). Now, the fact that paintings can have content but not necessarily denotation (for example, a painting of a unicorn does not refer to an actual object) makes things difficult for a pure denotation view: 'a picture must denote a man to represent him, but need not denote anything to be a man representation' (Goodman 1976, 25).

From considerations like these it follows that denotation is not even a *necessary* component in the explanation of representational content. Thus, the ability of something to represent, in the sense of having representational content, has nothing to do with its denotation. It is simply contingent, whether a painting has a denotation or not, although the painting has representational content in any case. Goodman is very explicit in some places in *Languages of Art* that (in cases, where denotation is not null) denotation alone does not suffice to make something a representation of something else. After all, an officer may use the paintings of a museum which has been occupied by his unit to denote the positions held by the enemy. In cases like that, no representation appears. In order to represent, the *content* of a painting has to relate to its denotation in some adequate way, such that 'what is denoted depends solely upon the pictorial properties of the symbol' (Goodman 1976, 42).

If representation could be constituted merely by denotation, we would get a concept of representation so weak that it would not be possible, for example, to explain interesting abilities of organisms like the successful homing behavior of desert ants by means of their ability to represent their own movements in their environment (see Gallistel 1990, 59f.). True, representations denote what they represent, but to understand how an organism performs well using a certain representational system we have to consider the specific *contents* of the representation and how they relate to its reference objects. Content is a necessary component of representation, and homomorphisms are necessary to explain this necessary component.

6. *The Copy Theory Objection*

Finally, I want to reject the common, but misguided view that the homomorphism theory is nothing more than a (more precise) version of the similarity theory of representation. My impression is that this misconception builds the core of the bundle of arguments against the theory that I have discussed before.

In section 3 it was noted that 'to be homomorphic to' is not a symmetric relation. Therefore, this relation is not a 'similarity'. Similarities are reflexive and symmetric. Since the notion of a 'similarity theory of representation' has never been made very clear, many (and often misplaced) connotations can invade that notion. One of the most popular connotations is that of a *copy theory*. The following short remarks are intended to show how the homomorphism theory is distinguished from a copy theory of representation:

It is a very common phenomenon that classes of objects be used to represent a given domain, for which no *analogous intrinsic relations* exist. For example, the natural numbers are represented by decimal symbols, although finite sequences of decimal symbols do not possess any intrinsic structure similar to the relation of addition defined on the natural numbers. In this case, we *extrinsically* endow the system of decimal symbols with the desired structure by imposing rules of calculation, in order to enable them to represent the addition structure of the natural numbers (such that the decimal representation of the sum of two natural numbers equals the sum of the decimal representations).

The representing objects very often are not simply there, but have to be constructed together with their relational structure for a certain representational use. In such cases, only when those objects have been constructed, the representandum can be identified with some part of the newly built object class (it is then said that the representandum has been *embedded* into the new domain). The representing domain is then not a ‘copy’ of the original domain in a twofold sense: firstly, because the representing domain is the result of a construction, and secondly, because the representing domain includes the original domain as a subdomain. For an example see Carnap’s construction of ‘quality classes’ as representations (logical reconstructions) of intuitive perceptual qualities (Carnap 1998, 98). In this case, neither the representandum, namely the qualities, nor the entities which represent them, are simply ‘given’. The qualities exist only in the sense that it is a common way of speaking to refer to ‘qualities’ as ‘parts’ of elementary experiences. This common way of speaking is, from the phenomenal perspective, incorrect. From such a perspective, elementary experiences do not have any parts. This incorrect mode of speech, according to Carnap, has to be replaced by a correct explication of ‘qualities’ by means of a logical reconstruction of quality classes. Thus, the quality classes themselves are clearly not ‘given’ before the representation procedure starts.

Carnap’s guideline for the construction of quality classes is that the classes have to fulfill a certain structural characterization: a quality class counts as an adequate reconstruction of a certain quality if the elementary experiences belonging to this quality class relate to each other in the same way as elementary experiences relate to each other when they ‘contain’ that quality. Quality classes exemplify qualities, that is, they are taken as models of the common expression ‘some two elementary experiences contain the same quality’. Whereas ‘copying’ of something does not lead to any new knowledge about that something, the construction of new objects —given the constraint of structure preservation, as explained above— may improve our knowledge with respect to both, its precision and its scope.

These cases demonstrate how representations can be generated by homomorphisms, although they are not ‘copies’. The decimal representation does not copy the additive structure of the natural numbers, and Carnap’s quality classes do not copy the qualities of our phenomenal experience. As I hope, the reader will be convinced by now that the most common conceptual objections against the structural concept of representation are misplaced. They seem to originate from a common source, the

copy theory misunderstanding. As this misunderstanding is swept away in the light of the notion of homomorphism, a more promising pursuit with regard to the structural theory is waiting: the pursuit of its empirical prospects.

REFERENCES

- Bailer-Jones, D. (2003). 'When Scientific Models Represent', *International Studies in the Philosophy of Science* 17: 59-74.
- Carnap, R. (1998). *Der logische Aufbau der Welt*. Hamburg: Meiner.
- Cummins, R. (1996). *Representations, Targets, and Attitudes*. Cambridge (MA): MIT Press.
- Dunn, M., and G. Hardegree (2001). *Algebraic Methods in Philosophical Logic*. Oxford: Clarendon Press.
- French, S. (2003). 'A Model-Theoretic Account of Representation (Or, I Don't Know Much About Art... But I Know It Involves Isomorphism)', *Philosophy of Science* 70 (Proceedings): 1472-1483.
- , and N. da Costa (2003). *Science and partial Truth*. Oxford: Oxford University Press.
- Gärdenfors, P. (2000). *Conceptual Spaces*. Cambridge (MA): MIT Press.
- Gallistel, C. R. (1990). *The Organization of Learning*. Cambridge (MA): MIT Press.
- Goldstein, E. B. (1996). *Wahrnehmungspsychologie*. Heidelberg: Spektrum.
- Goodman, N. (1976). *Languages of Art*. 2nd ed., Oxford: Oxford University Press.
- Grush, R. (1995). *Emulation and Cognition*. Ph.D. Dissertation. San Diego: University of California.
- Hughes, R. I. G. (1997). 'Models and Representation', *Philosophy of Science* 64 (Proceedings): S325-S336.
- Ibarra, A., and T. Mormann (2000). 'Una teoría combinatoria de las teorías científicas', *Crítica* 32 (95): 3-46.
- Mundy, B. (1986). 'On the General Theory of Meaningful Representation', *Synthese* 67: 391-437.
- Scholz, O. (1991). *Bild, Darstellung, Zeichen*. Freiburg: Alber.
- Suárez, M. (2003). 'Scientific representation: Against similarity and isomorphism', *International Studies in the Philosophy of Science* 17, 3: 225-244.
- (2004). 'An Inferential Conception of Scientific Representation', *Philosophy of Science* 71 (Proceedings), 5: 767-779.
- Swoyer, C. (1991). 'Structural Representation and Surrogate Reasoning', *Synthese* 87: 449-508.
- Watson, R. A. (1995). *Representational Ideas. From Plato to Patricia Churchland*. Dordrecht: Kluwer Academic Publishers.

Andreas BARTELS: Ph.D. 1984 Giessen: "Causality Violation in General Relativity". Scientific interests: Philosophy of Physics, Epistemology, Theory of Concepts, Representation. Chair for Philosophy of Science at the University of Bonn, Germany (since 2000). Project on "Knowing that and Knowing how" (funded by VolkswagenStiftung). Books: *Bedeutung und Begriffsgeschichte* (1994), *Grundprobleme der modernen Naturphilosophie* (1996), *Strukturelle Repräsentation* (2005).

ADDRESS: Institut für Philosophie, Universität Bonn, Am Hof 1, D-53113 Bonn, Germany. E-mail: andreas.bartels@uni-bonn.de.

Scientific Theories as Intervening Representations¹

Andoni IBARRA, Thomas MORMANN

BIBLID [0495-4548 (2006) 21: 55; pp. 21-38]

ABSTRACT: In this paper some classical representational ideas of Hertz and Duhem are used to show how the dichotomy between representation and intervention can be overcome. More precisely, scientific theories are reconstructed as complex networks of intervening representations (or representational interventions). The formal apparatus developed is applied to elucidate various theoretical and practical aspects of the *in vivo/in vitro* problem of biochemistry. Moreover, adjoint situations (Galois connections) are used to explain the relation between empirical facts and theoretical laws in a new way.

Key words: Representation, adjoint situations, *in vitro/in vivo* problem, Hertz, Duhem.

1. Introduction

The concept of representation has not yet been secured on the agenda of philosophy of science. Some philosophers flatly deny that it could be of any use in epistemology or philosophy of science. Instead, they claim, the concept of representation leads us into a hopeless maze of pseudo-questions without answers. This is the case of Rorty and his antirepresentationalist followers. According to them, epistemology based on the notions of negotiation and interpretation should replace epistemological accounts based on ‘representation’. In this paper, we will not address this kind of radical anti-representationalism. But suffice to say, it is based on a rather primitive conception of representation identifying representation with some kind of copying or mirroring.

In this paper we want to elaborate some classical representational ideas of Hertz and Duhem in order to show that a diagrammatical or combinatorial account of representations can be useful for elucidating the role of representations in describing the practice of representational reasoning in science.

The outline of this paper is as follows: in section 2, we outline some ideas of Hertz and Duhem concerning the structure of scientific reasoning that can be used to understand how representations in science work. More precisely, following Hertz the idea of a commutative diagram of interconnected representations is introduced, and Duhem’s account of empirical theories will lead us to the idea that the theoretical and the empirical are correlated in a so called adjoint situation. In section 3, the rudiments of a combinatorial theory of representations are introduced, and are put to use in section 4 for the representational elucidation of the *in vitro/in vivo* problem in biochemistry. In section 5, it is shown that Duhem’s account of an empirical theory as a cor-

¹ We would like to thank two anonymous referees for their detailed and penetrating criticisms that helped us to correct some major blunders in the first version of this paper. Further we’d like to express our sincere gratitude to the guest editor José Antonio Díez, who pointed out some conceptual obscurities and infelicitous formulations in the original text. Of course, we are responsible for all remaining errors.



relation of symbolical and empirical facts leads to the conception of an empirical theory as a Galois connection (or, more generally, an adjoint situation) in the sense of mathematical category theory. We close with some general remarks on the role of representational concepts in philosophy of science.

2. Classical Ideas of Representations

Let us start with some basic ideas on scientific representations put forward by the classical philosopher-scientists Hertz and Duhem. These ideas naturally lead towards an interesting conception of scientific theories as representations.

As our first classical intuition pump for the development of a comprehensive account of representation, we take Hertz's well known 'symbolical account' put forward in his *The Principles of Mechanics presented in a New Form* (Hertz 1894) where he described the general procedure of scientific representations as follows:

We form for ourselves images or symbols of external objects; and the form which we give them is such that the necessary consequents of the images in thought are always the images of the necessary consequents in nature of the things pictured. In order that this requirement may be satisfied, there must be certain conformity between nature and our thought. Experience teaches us that the requirement can be satisfied, and hence that such a conformity does in fact exist. [...]

The images, which we may form of things, are not determined without ambiguity by the requirement that the consequents of the images must be the images of the consequents. Various images of the same objects are possible, and the images may differ in various respects. [...]

Of two images of equal distinctness the more appropriate is the one which contains, in addition to the essential characteristics, the smaller number of superfluous or empty relations, —the simpler of the two. Empty relations cannot be altogether avoided: they enter into the images because they are simply images ... produced by our mind and necessarily affected by the characteristics of its mode of portrayal. (Hertz 1894, pp. 1f.)

We propose to translate Hertz's informal description of the representational activity of science in a diagrammatical language as follows: let the set of "external objects" be denoted by E , and denote the set of "images" by S . The following diagram may be used to capture the essential structure of Hertz's account:

$$\begin{array}{ccc}
 E & \xrightarrow{t} & S \\
 f \downarrow & & \downarrow g \\
 E & \xrightarrow{t} & S
 \end{array}$$

The details are as follows: the horizontal arrow t corresponds to Hertz's formation of mental images. More precisely, if $e \in E$ is an external object, $t(e) \in S$ is the image corresponding to it. In other words, $t(e)$ may be considered as the theoretical counterpart of e . The left vertical arrow f in Hertz's diagram is to be conceived as a process or an experiment that 'necessarily' brings about the external fact that e is changed to another

external fact $f(e) \in E$. In Hertz's terms, $f(e)$ is the 'necessary consequent' of e . Analogously, the vertical arrow g on the right may be interpreted as a mathematical calculation or a logical argument that leads from a 'symbol' $s \in S$ to another symbol $g(s)$. It is to be interpreted as the result or the conclusion of the symbolical transaction g . In Hertz's terms, $g(s)$ is the 'necessary consequent' of s . These ingredients of Hertz's diagram are of course not independent of each other; rather, as is informally stated in his *Principles*, they form a commutative diagram in 'that necessary consequents of the images in thought are always the images of the necessary consequents in nature', which in our diagrammatical language just amounts to the commutativity of the diagram:

(2.1) *Commutativity of Hertz's Diagram.* Assume t, f , and g as characterized above. They are assumed to satisfy the following concatenation law:

$$g \bullet t = t \bullet f$$

This equation is to be interpreted as follows: If we start with an empirical fact e in the left upper corner of the Hertz diagram, translate it to its theoretical counterpart $t(e)$, and use $t(e)$ as the input for a calculation or a logical argument that leads to $g \bullet t(e)$, then this outcome is the same as if we had submitted the empirical fact e to an experimental transformation f arriving at $f(e)$, and translated this experimental fact $f(e)$ by t finally yielding $t \bullet f(e) = g \bullet t(e)$. In other words, the two paths in Hertz's diagram are strictly equivalent in that they may be considered as paths that lead to one and the same destination. As an elementary example consider e to be some chemical substance that is submitted to a certain chemical experiment f which, say, oxidizes e thereby yielding as outcome another chemical substance $f(e)$. For this transaction a chemical theory has to provide a chemical formula $t(e)$ for e , and a theoretical transformation $g(t(e))$ of $t(e)$ such that $t(f(e)) = g(t(e))$. As is emphasized by Hertz, given E there may be different 'symbolic completions' S, S' . The choice between them is a pragmatic matter of simplicity and local usefulness. It may be that for different purposes different 'images' may be appropriate (cf. Hertz 1894, p. 3).

Second, let us come to Duhem's contribution to a modern representational account of scientific theorizing, which is found in his classic *The Aim and Structure of Physical Theory* (Duhem 1906). At various occasions in his *opus magnum* he asserts that scientific theories are to be conceived as representations. More precisely, he considers a physical theory 'as an economical representation' that

establishes an order and a classification among [the experimental laws]. It brings some laws together, closely arranged in the same group; it separates some others by placing them into two groups very far apart. Theory gives, so to speak, the table of contents and the chapter headings under which the science to be studied will be methodologically divided. (Duhem 1906, pp. 23f).

Later he goes on to explain this 'representation' as a correspondence between 'practical facts' and 'theoretical' or 'symbolical facts'. It is certainly not too far fetched to consider Duhem's account as presented up to now as just another version of Hertz's structural approach. But there is one feature in Duhem's representational ap-

proach that is novel and not present in Hertz. In describing a physical theory as a correspondence between practical and symbolical facts he insists that

a symbolic formula ... can be translated into concrete facts in an infinity of different ways, *because all these disparate facts admit the same theoretical interpretation.* (*Ibid.*, p. 150)

And, in an analogous vein:

The same practical fact may correspond to an infinity of logically incompatible theoretical facts; the same group of concrete facts may be made to correspond in general not with a single symbolic judgment but with an infinity of judgments different from one another and logically in contradiction with one another. (*Ibid.*, p. 152)

Duhem's account is rather informal, and he is not very clear about what is to be understood by 'theoretical fact'. In particular, one should not interpret him as conceiving a 'theoretical fact' as a fact 'belonging' to a specific theory. Rather, the most appropriate interpretation of Duhemian theoretical facts is to take a theoretical fact as one that asserts a physical state of affairs in precise mathematical terms, as is explained by Duhem. A typical example of a theoretical fact (or statement) is the following: 'An increased pressure of 100 atmospheres causes the electromotive force of a given gas battery to increase by 0.0844 volts.' (*Ibid.*, p. 152) Other 'logically incompatible' theoretical statements would be obtained by replacing '0.0844' by '0.0845' or '0.0846'. Hence, Duhem's account of an empirical theory can be formulated in relational terms as follows:

(2.2) *Duhem's Relational Account of Empirical Theories.* Denote the class of symbolic facts by S and the class of practical or empirical facts by E . Then a theory T is to be conceived as a relation

$$T \subseteq E \times S.$$

If $(e, s) \in T$ then this is to be interpreted as the empirical fact that e is related to s , or, to put it the other way round, that the symbolic fact s is related to the empirical fact e .

It is important to note that Duhem insisted that this relation is multi-valued: to a single e there may correspond many symbolic facts s , and, vice versa, to a single s , there may correspond many empirical facts e . This double ambiguity of the relation between empirical and symbolical facts is characteristic of Duhem's account and has no counterpart in Hertz's approach. As we shall show in the next section, this feature may be combined with the representational insights of Hertz to yield a complex representational account of empirical theories.

3. Representational Combinatorics

Following Hertz and Duhem in conceiving the practice of science as engaged in producing and manipulating representations of various kinds, the impression that comes to mind is that scientific representations do not live in isolation, rather they may be combined and concatenated in various ways (Ibarra, Mormann 2000). Hence, investigating these combinatorial aspects of representations is a central task of a general theory of representation (Ibarra, Mormann 1997 a, b).

Regardless of what kind of representations we consider, they are not unconnected with each other, rather, they form a representational network. One and the same entity A may be represented by several different entities B, C, D etc. such that we have representations $A \xrightarrow{r} B, A \xrightarrow{s} C, A \xrightarrow{t} D$, etc. On the other hand, it may happen that one and the same entity E appears as the representative of several different entities A, B, C etc. That is to say we have representations $A \longrightarrow E, B \longrightarrow E, C \longrightarrow E$. Furthermore, it can be the case that representations such as $A \xrightarrow{r} B$ and $B \xrightarrow{s} C$ are concatenated yielding an indirect or combined representation $A \xrightarrow{s \cdot r} C$.

As the result of these considerations, we can see that any theory of representations should comprise a combinatorial part, which describes the various possibilities of combinations and iterations of representations. In the following we shall assume that this combination or concatenation of representations is associative, i.e. representations f, g and h , which 'match', satisfy the following law of associativity:

$$(3.1) \quad f \cdot (g \cdot h) = (f \cdot g) \cdot h.$$

The combination or iteration of representations is of utmost importance for the practice of science. For instance, in the standard representational theory of measurement the numerical measurement of an empirical domain D is conceptualized as a representation $r: D \longrightarrow \mathfrak{R}$ of D into the real numbers \mathfrak{R} . This is a rather idealized description. Actually, by a closer inspection the representation $D \xrightarrow{r} \mathfrak{R}$ should be regarded as a more or less extended chain of representations

$$(3.2) \quad D \longrightarrow E \longrightarrow F \longrightarrow \dots \longrightarrow \mathfrak{R}.$$

In most cases, numerical or, more generally, mathematical representations of empirical data cannot be 'read off' directly; usually they have to be considered as constructs which have been built by a more or less complicated constructional process. The long way from data to theory shows that the standard dichotomy is, at best, a very idealized picture. Dealing with an example from general relativity theory, Laymon gives a detailed account of the 'long contrafactual path from data to theory' (cf. Laymon 1982). Other examples of complex 'long distance' representations are discussed in detail in Latour (1999): Latour tells us in detail the long story from raw findings to theoretically digestible data in the case of 'botanical pedology' (*ibid.*, chapter 2). Notwithstanding important differences, all these accounts rely—implicitly or explicitly—on what may be called a combinatorics of representations.

The combination of representations is not restricted, however, to linear combinations. As will be shown by the *in vivo/in vitro* example of biochemistry, the point of the combinatorial account of representations only comes to the fore if we do not restrict

our attention to linear chains of representations but, instead, also take into account non-linear net-like configurations of representations.

The importance of representational nets or diagrams is evidenced by the fact that in the last forty years or so mathematics (and parts of other sciences as well) has been successfully reformulated in terms of representational networks. Here we refer, of course, to the mathematical theory of categories founded by Eilenberg and Mac Lane in the forties and presented for the general scientific public in books such as Mac Lane's *Mathematics – Form and Function* (1986) or Lawvere and Schanuel's *Conceptual Mathematics – A First Introduction to Categories* (1996). In category theory, representations appear under the names morphisms, functors, and natural transformations. In the last decades it has been shown that not only the bulk of mathematics can be reconstructed in these terms, but also that this representational reconstruction has led to new and fruitful lines of mathematical research. We take this fact, together with the representational ideas of Hertz and Duhem as evidence that combinations of various kinds of representations play an indispensable role for a representational theory of scientific knowledge. This claim is substantiated in the next section in which we propose to study in some detail various combinations of representations that arise from the so-called *in vitro/in vivo* problem in biochemistry.

4. *A Representational Account of the In Vivo/In Vitro Problem*

In this section we are going to apply the formal apparatus sketched so far to a specific problem of a scientific discipline that up to now has not received too much attention from philosophy of science, to wit, the so called '*in vitro/in vivo*' problem of biochemistry (cf. Strand, Fjelland, and Flatmark 1996 and Strand 1999). For the information on biochemical matters we heavily rely on these papers. Our purpose is to show that the rudiments of a theory of meaningful representations set out in the previous sections may be used to elucidate the problems of the representational practice biochemistry have to cope with. We chose the approach of Strand *et al.* as our starting point since it seemed to us particularly well suited for our purposes: on the one hand, it is sufficiently complex to require the employment of some non-trivial representational tools; on the other hand, it is conceptually not too complex as to be inaccessible for non-experts in biochemistry.

First, let us recall the basic ingredients of the *in vitro/in vivo* problem as it presents itself in biochemistry. The first point to note is that although biochemistry may be defined as 'the field of science concerned with the chemical substances and processes that occur in plants, animals, and microorganisms' it would be misleading to assume that 'biochemists study processes that occur in living organisms' (cf. Strand 1999, p. 273). The reason is that normally

it is impossible to perform a chemical analysis of an intact organism. A biochemical analysis is typically preceded by an isolation procedure, in which the organism of interest is disrupted and a specific component of it is isolated. To put another way, almost all biochemical evidence is obtained *in vitro* under artificial experimental conditions. ... [Nevertheless] biochemists are concerned with the chemistry of the living organism, *in vivo*. (Strand 1999, p. 273)

Hence one may even assert:

It would be wrong to say biochemists *observe* or *describe* or *study* processes that occur in living organisms, because they very rarely do so. Normally, it is impossible to perform a chemical analysis of an intact organism. (*Ibid.*, p. 273).

Almost all biochemical evidence is obtained *in vitro*, under artificial experimental conditions.

An *in vivo* system is a biologically interesting but experimentally inaccessible system, and the corresponding *in vitro* system as a related accessible, but biologically less interesting system. Although analogous situations also occur in other sciences, the difference between *in vitro* and *in vivo* systems is particularly striking in biochemistry. Now we may define the *in vitro/in vivo problem* (or the IVIV-problem henceforth) as the problem of justifying knowledge claims about *in vivo* systems on the basis of evidence obtained in ‘corresponding’ *in vitro* systems. Or, on a more descriptive level, the IVIV-problem may be said to be the problem of describing as clearly as possible the various methods used by biochemists to extract the information on *in vivo* systems they are seeking from the evidence they have obtained from *in vitro* systems.

One has to note that the IVIV distinction is a relative distinction. That is to say, in one context a system may play the role of the *in vitro* part, in another context the same system may be considered as the *in vivo* component. Of course, one may say that in many sciences one finds analogous distinctions to that of the IVIV distinction in biochemistry. Nevertheless, the case of biochemistry is special since the IVIV is thus central for this discipline, as is convincingly pointed out by Strand (1999, pp. 274f). His discussion may be summed up in the contention that the concept of artifact is central in any biochemical discussion. A biochemical artifact is a chemical reaction that occurs between biomolecules *in vitro*, but not *in vivo*. Now, the problem of artifacts is a central problem of meaningful representations everywhere. Given a representation $A \xrightarrow{r} B$ the problem arises to interpret elements and relations defined on B in terms of A , for instance, if $r(a) = r(a')$ one may ask if this identity on the representing domain B may be pulled back to A , i.e., one asks if it is possible to infer $a = a'$. Consider first the special case that r is a function. Then, of course, this inference is not valid in general. It is only admissible to infer from $r(a) \neq r(a')$ that $a \neq a'$. Now let us consider the general case that the representation r is any relation between A and B . Denote the power set of B by PB . Then, committing an innocent abuse of language, r may be conceived as a function $A \xrightarrow{r} PB$ defined by $r(a) := \{b; (a, b) \in r\}$. In the same vein as above one can infer from $r(a) \neq r(b)$ that $a \neq b$. In other words, the inequality on B (or PB) can be pulled back to an inequality on A . Of course, the problem of artifacts is not restricted to this kind of artifacts. Other, more complicated relations on B such as $R(f(a_1), \dots, R(f(a_n)))$ may be considered and tested for their A -meaningfulness.

We take the fact that the problem of artifacts can be naturally couched in representational terms as evidence that the IVIV problem should be treated in terms of a theory of meaningful representations. Thus we propose to conceive the relation between

an *in vivo* system S and a corresponding *in vitro* system S^* as a representational relation $S \xrightarrow{d} S^*$ contending that the *in vivo* system S is represented by the *in vitro* system S^* .

First, it should be noted that this representation is a material long distance representation par excellence: Usually the representing system S^* is obtained from S by a variety of massive, often destructive interventions of various kinds (cf. Strand *et al.* 1996, Strand 1999). The representing system S^* is far from being similar to S , and it is neither natural nor necessary to represent S by S^* . There may be many other ways of representing S by other S^* , S^{**} , ... depending on the representational interests and capacities of those who are engaged in the construction of these *intervening* representations. Thus, as the first outcome of considering the IVIV problem in biochemistry we contend that the dichotomy between representing and intervening put forward by some philosophers such as Hacking is pointless in the case of biochemistry, and, regarding biochemistry as a paradigmatic case for science in general, for other sciences as well (cf. Hacking 1983).

As lucidly explained by Strand *et al.*, there is much more in the IVIV problem than the statement that it gives rise to an intervening representation $S \longrightarrow S^*$. To deal with these more fine-grained aspects of the IVIV problem, let us introduce the following terminological conventions: properties, objects, relations, procedures etc. belonging to the realm of *in vivo* systems are denoted by E, F, a, b, R, p, \dots , while the corresponding properties, objects, etc. belonging to the *in vitro* realm are denoted by $E^*, F^*, a^*, b^*, \dots$. Our first purpose is to show that IVIV problems give rise in a natural way to a plurality of Hertz's diagrams. Given systems S and S^* , and important task of the biochemist's work is to study how these systems behave under certain perturbations p and p^* . Here, a perturbation p of S may be considered as a map: $S \xrightarrow{p} S$. More precisely, $p(s)$ is to be understood that for $s \in S$ the state $p(s) \in S$ is the state that resulted from s when submitted to the perturbation p . Analogously for *in vitro* states S^* and *in vitro* perturbations p^* : $S^* \longrightarrow S^*$. Then the systems and perturbations S, p, S^*, p^* may be said to be optimally correlated if the following Hertz diagram commutes:

$$\begin{array}{ccc}
 S & \xrightarrow{d} & S^* \\
 p \downarrow & & \downarrow p^* \\
 S & \xrightarrow{d} & S^*
 \end{array}
 \quad p^* \bullet d = d \bullet p.$$

By definition, an artifact is an *in vitro* perturbation $d(s) \neq p^*(d(s))$ such that $s = p(s)$. If the Hertz diagram commutes, artifacts can be shown not to exist: Assume $d(s) \neq p^*(d(s))$ and $s = p(s)$. From Hertz we get $p^*(d(s)) = d(p(s))$. Hence we get the following proposition:

Proposition 1. If Hertz commutes, then there are no artifacts.

In a similar vein, one obtains that the non-existence of artifacts implies that the Hertz diagram commutes for states s that are invariant under the perturbation p , i.e., states for which $s = p(s)$:

Proposition 2. If s is invariant under p AND there are no artifacts, then HERTZ commutes for s .

Proof. Assume $s = p(s)$. Then $d(s) = d(p(s))$. Assume that HERTZ does not commute for s . That is to say $p^*(d(s)) \neq d(p(s))$. Then $p^*(d(s)) \neq d(s)$. Since there are no artifacts one infers $s \neq p(s)$. This is a contradiction. ■

In sum, the diagrammatically natural requirement that Hertz diagrams commute is a bit stronger than the claim that no artifacts exist. The existence of artifacts is, however, not the only problem that may arise when studying the relation between *in vivo* and *in vitro* systems. It may well happen that the combination of *in vitro* perturbation $p^*: S^* \longrightarrow S^*$ and the intervening representation $d: S \longrightarrow S^*$ are jointly too invasive and too coarse, such that a salient *in vivo* perturbation p fails to be detected by them. This is the case if it happens that $s \neq p(s)$ but $d(s) = p^*(d(s))$. This may be called an artificial null effect. Artificial null effects and the commuting of the Hertz diagram are related as follows:

Proposition 3. If the Hertz diagram commutes and the representation $d: S \longrightarrow S^*$ is mono, i.e., $d(a) = d(b)$ implies $a = b$, then no artificial null effects occur. ■

In this implication, the second clause of the antecedent is clearly necessary. This may be more conspicuously expressed by contraposition:

Proposition 4. If artificial null effects occur, then either the Hertz diagram does not commute or the IVIV representation $d: S \longrightarrow S^*$ is not mono. ■

One may ask whether the converse holds: If no artificial null effects occur, does the Hertz diagram commute and is d mono? As is easily checked by examples, this is not the case. In other words, the conjunctive assumption that the Hertz diagram is commutative and the IVIV representation d is mono is strictly stronger than the non-existence of artificial null effects.

As has been pointed by Strand *et al.*, the IVIV problem is not completely described by a Hertz diagram connecting an *in vivo* systems S and an *in vitro* systems S^* . Usually these systems are accompanied by what may be called their model systems M and M^* respectively. That is to say, for the *in vivo* system S there is a theoretical (or maybe sometimes a computer) model M , and for the *in vitro* system S^* there is a theoretical (computer model) model M^* . Then it is natural to assume that M is an appropriate representation of S , and M^* is an appropriate representation of S^* . These may be ex-

plicated by the assumption that the representations $S \xrightarrow{t} M$ and $S^* \xrightarrow{t} M^*$ have Hertz diagrams of the following kind:

$$(4.1) \quad \begin{array}{ccc} S & \xrightarrow{t} & M \\ \downarrow & & \downarrow \\ S & \xrightarrow{t} & M \end{array} \quad \begin{array}{ccc} S^* & \xrightarrow{t} & M^* \\ \downarrow & & \downarrow \\ S^* & \xrightarrow{t} & M^* \end{array}$$

For each of these diagrams one may study the various ways in which artifacts may influence the reliability of surrogative reasoning dealing with M , S^* , and M^* and finally bound to obtain information about the *in vivo* system S .

For dealing in a reasonable way with problems of this kind it is not sufficient, however, to assume that Hertz diagrams for (S, S^*) , (S, M) , and (S^*, M^*) exist. One has to assume the existence of a further ‘purely theoretical’ Hertz diagram for (M, M^*) such that the following ‘3-dimensional’ or ‘cubical’ diagram commutes:

$$(4.2) \quad \begin{array}{ccccc} & & S^* & \xrightarrow{t^*} & M^* \\ & \nearrow & \vdots & \nearrow & \downarrow \\ S & \xrightarrow{t} & M & \nearrow & \\ \downarrow & & \downarrow & & \downarrow \\ S & \xrightarrow{t} & M & \nearrow & \\ & \searrow & S^* & \xrightarrow{t^*} & M^* \\ & \downarrow & \vdots & \searrow & \\ & & S & \xrightarrow{t} & M \end{array}$$

Of course, it can hardly be expected that in reality the cube (4.2) is fully commutative. Rather, there will exist various sources of non-commutativity, which show that the various kinds of systems and models only match approximately. Nevertheless, the cube presentation (4.2) may be useful as an idealized model to spot where precisely commutativity and thereby the validity of surrogative reasoning via models and systems of various kinds may fail.

Let us consider a particularly simple theoretical model of (*in vivo* or *in vitro*) systems, which, at first, may not appear as models at all. Assume that for a given system S the possible states s of S may have certain properties. This assumption may be cast in a representational framework by stipulating that there is a map $F: S \longrightarrow C$, C being a structure whose elements are to be interpreted as properties belonging to a certain property type. In other words, $F(s)$, $s \in S$, is to be conceived as the assertion that the state s has the property $F(s)$. Given a perturbation $p: S \longrightarrow S$ one may ask, if F is in-

variant with respect to p , i.e., if $F(p(s)) = F(s)$, or not. Analogously, for *in vitro* properties $F^*: S^* \longrightarrow C^*$ a corresponding *in vitro* perturbation $p^*: S^* \longrightarrow S^*$ is defined that may or may not be invariant under the *in vitro* property $F^*: S^* \longrightarrow C^*$. The properties F and F^* are correlated by d and d_c iff there is a commutative diagram of the following kind:

$$\begin{array}{ccc}
 \begin{array}{c} \curvearrowright \\ S \end{array} & \xrightarrow{F} & C \\
 \downarrow d & & \downarrow d_c \\
 \begin{array}{c} \downarrow \\ S^* \end{array} & \xrightarrow{F^*} & C^* \\
 & \begin{array}{c} \curvearrowleft \\ \end{array} &
 \end{array}$$

This diagram describes the (ideal) relation between *in vivo* properties F and *in vitro* properties F^* . From now on, let us assume that F and F^* are such that there exist d and d_c so that the diagram commutes. This means that F and F^* are reasonably correlated with each other. This is to ensure that assertions dealing with F^* may be possibly translated into assertions dealing with F , that is to say that F and F^* can be correlated by surrogate reasoning. On this non-trivial assumption about F and F^* is based the entire *in vivo/in vitro* argumentation.

Usually, the domain of values C of a property F is not just a set, but has some structure. For instance, often C is assumed to be endowed with an order relation \leq . Then we may define an order relation on S by pulling back the order defined on C by the following definition:

$$s \leq s' := F(s) \leq F(s').$$

In an analogous way the state space S^* of an *in vitro* system may be endowed with an order via a map $S^* \xrightarrow{F^*} C^*$ of S^* into an ordered property space C^* . The *in vivo* property F is stable under the *in vivo* perturbation $p: S \longrightarrow S$ iff $s \leq s'$ implies $p(s) \leq p(s')$, i.e., iff $F(s) \leq F(s') \Rightarrow F(p(s)) \leq F(p(s'))$. Analogously for *in vitro* perturbation p^* and an *in vitro* feature F^* . Of course, the two properties F and F^* and the perturbations p and p^* should not be unrelated to each other. Rather, the *in vivo* property F and the *in vitro* property F^* should obey the following relation:

$$F(p(s)) \leq F(p(s')) \Rightarrow F^*(p^*(d(s))) \leq F^*(p^*(d(s'))).$$

In this case, from the accessible *in vitro* relation $\text{NOT}(F^*(p^*(d(s))) \leq F^*(p^*(d(s'))))$ one can infer $\text{NOT}(F(p(s)) \leq F(p(s')))$.

Discussing the structural features of the IVIV problem shows that in the biochemical practice the concepts of representation and intervention are intimately related. More precisely, they are correlated by a variety of commutative diagrams that combine *in vivo* systems, *in vitro* systems, *in vivo* models, and *in vitro* models by a com-

plex net of representational and intervening links. The basic building block of this net, which intertwines theoretical representations and practical interventions are various kinds of Hertz diagrams. Thus, taking the IVIV problem in biochemistry as paradigmatical for empirical theories in general we contend that representations and interventions should be treated together, since both may be characterized as moves in the complex network of an empirical theory.

The IVIV problem of biochemistry is particularly interesting for a representational philosophy of science as it shows the necessity of considering iterations and combinations of various kinds of interventions and representations. The language of representational diagrams is particularly apt for dealing with the various kinds of connections. We think that the opposition between the representative and the performative perspective in philosophy of science is an artifact of a misinformed philosophy of science. One does not have to choose between them. Indeed, in some sense, every representation has an interventional aspect, at least indirectly, and every intervention leads to a representation.

5. *Adjoint Situations*

In this section we are going to show that Duhem's relational account of theories that conceives a theory T as a relation $T \subseteq S \times E$ between symbolic and empirical facts may be elucidated by using so called Galois connections or, more generally, adjoint situations in the sense of category theory. This part of the paper is the most speculative one, and some readers may object that we introduce a heavy formal apparatus without real justification. Thus the following preliminary remark may be in order. Our point is this: conceiving an empirical theory as a certain relation between empirical and theoretical facts seems to us quite a natural and intuitive approach. Otherwise Duhem, who certainly was not interested in formal technicalities, would not have endorsed it. Now, as soon as a theory is given as a relation $T \subseteq S \times E$, the whole apparatus of Galois relations is available. One may even say that a Galois relation between PS and PE is nothing but a relation. Since Galois connections have turned out to be a useful tool in the study of binary relations in mathematics, computer science and elsewhere. Hence, one may suspect that they could do some useful work in formal philosophy of science as well. This conjecture is further supported by the fact that Galois connections are just a very special case of adjoint situations that may be characterized as *the* fundamental concept of category theory. Hence, there is some hope that these conceptual tools have some applications in philosophy of science as well.

By conceiving a theory as a relation $T \subseteq E \times S$ of empirical and symbolical facts in the sense of Duhem's *The Aim and Structure of Physical Theory*, it is not claimed, of course, that any relation $X \subseteq S \times E$ counts as a genuine theory. There are countless relations between the two classes of facts that make no sense at all. Further restrictions will have to be imposed on T in order that T can be acknowledged as a genuine theory. As will be shown later, for this task the representational ideas of Hertz turn out to be useful.

For the moment we only want to emphasize Duhem's main point, to wit, that for any given empirical fact $f \in E$ there may be many symbolic facts $s \in S$ such that f and s are theoretically correlated, i.e., that $(f, s) \in T$, and that vice versa for any $s \in S$ there may be many empirical facts $f \in E$ such that $(f, s) \in T$ (cf. Duhem 1906, pp. 152ff). Formally, this means that $T \subseteq E \times S$ is a relation and not a function.

This multivalued correlation between empirical and symbolical facts renders it plausible that a single fact, be it symbolical or empirical, hardly makes sense as such. That is to say, a single $s \in S$ or $f \in E$ is an object that in real science hardly occurs. Rather, what shows up in the practice of real science are *clusters* or *complexes* of empirical and theoretical facts. Thus, we propose to consider appropriate *sets* $A \subseteq S$ and $B \subseteq E$ as the real building blocks of scientific theories; single empirical facts $f \in E$ or symbolic facts $s \in S$ are auxiliary concepts introduced for methodological reasons. Replacing elements by subsets in this way is a natural generalization in so far as the 'elementary' facts of type s and f may be considered as special cases of facts of type A and B by identifying s and f with their singletons $\{s\}$ and $\{f\}$. This technical move from elementary facts to subsets of elementary facts resembles the approach Duhem's Austrian colleague Ernst Mach proposed long time ago: according to Mach, it was the task of science to describe the functional relations of appropriate complexes or clusters of elements in the most economical way possible. In any case, the move from elements to subsets facilitates to get started the formal apparatus we are going to apply in order to elucidate Duhem's relational account of scientific theories. After these preparatory remarks we are now ready to set up the formal apparatus we need in order to cast Duhem's relational account of empirical theories in the framework of Galois connections. First, let us deal with the necessary technicalities.

Denote by PS and PE the power sets of S and E , respectively. For the moment, let us assume that PS and PE are endowed with their natural (set-theoretical) order structures (PS, \subseteq) and (PE, \subseteq) . A theory $T \subseteq E \times S$ gives rise to order-preserving maps between PS and PE by the following recipe:

(5.1) *Proposition.* Let $T \subseteq E \times S$ be a theory. Define maps $PE \xrightarrow{t} PS$ and $PS \xrightarrow{e} PE$ by:

- (a) For $Y \in PE$ define $e(Y)$ by $e(Y) := \{s; \exists y (y \in Y \text{ AND } (y, s) \in T)\}$
- (b) For $X \in PS$ define $t(X)$ by $t(X) := \{y; (e(\{y\}) \subseteq X)\}$.

Then the maps e and t are order preserving.

Proof. Check the definitions of e and t .

Obviously, e and t are not unrelated to each other. Indeed, it can be shown that t is completely determined by e , and vice versa. Actually, much more is true, as is shown by the following proposition:

(5.2) *Proposition.* Let e and t be defined as above. Then for all $X \subseteq S$ and $Y \subseteq E$ the following holds:

$$X \subseteq t(Y) \text{ IFF } e(X) \subseteq Y.$$

In technical jargon, the ordered pair (t, e) is called a *Galois connection* between the order structures PS and PE (cf. Gierz *et al.* 2003). More precisely, t is called the upper (or right) adjoint, and e is called the lower (or left) adjoint. One should note that a Galois connection (t, e) is *not* a symmetric notion, i.e., if (t, e) is a Galois connection, usually (e, t) fails to be a Galois connection. The difference between upper and lower adjoint is reflected in the notational convention that t as the upper adjoint is on the right or ‘upper’ side of \leq , while e as the lower adjoint is on the ‘lower’ side of the order relation \leq . This asymmetry is essential in the following to set up an asymmetric relation between the domain of empirical facts E and the domain of symbolical facts S .

Proof (5.2). The proof naturally splits into two parts: (i) assume $X \subseteq t(Y)$ and $z \in e(X)$. Then one has to show $z \in Y$. By definition of $e(X)$ there is an $s \in X$ with $(s, z) \in T$. That is to say $z \in e(s)$. By presupposition $s \in t(Y)$. This means $e(s) \subseteq Y$, and therefore $z \in Y$; (ii) Assume $e(X) \subseteq Y$ and $s \in X$. One has to show $s \in t(Y)$. But $e(s) \subseteq e(X) \subseteq Y$, and this just means $s \in t(Y)$.

(5.3) *Corollary.* The map $PS \xrightarrow{e \bullet t} PS$ is a kernel operator, i.e., $e \bullet t(X) \subseteq X$, for all $X \subseteq S$, and the map $PE \xrightarrow{t \bullet e} PE$ is a closure operator, i.e. $Y \subseteq t \bullet e(Y)$ for all $Y \subseteq E$.

After having presented these rudiments of the theory of Galois connections, let us start now with the task of elucidating the intuitive meaning of this gadget. This amounts to an interpretation of the components t and e , which form the Galois connection (t, e) , and an explanation of their most important properties in informal terms of philosophy of science.

For this task it is expedient to start with the map $e: PS \longrightarrow PE$. Recall that subsets $X \subseteq S$ and subsets $Y \subseteq E$ are to be interpreted as symbolic (theoretical) and empirical facts. By definition $e(X)$ is the collection of all ‘atomic’ empirical facts z that are empirically correlated to at least one ‘atomic’ symbolic fact $s \in X$. This may be interpreted as that the empirical fact $e(X)$ provides an empirical realization of X in a broad sense, i.e., it may be that the empirical facts z realizing the symbolic facts of X may have theoretical correlates s that do not belong to X but at least X is covered by the empirical facts of $e(X)$ in the sense that $t(e(X)) \supseteq X$.

Analogously, the map t may be interpreted as a recipe to translate an empirical fact $Y \subseteq E$ into a related theoretical fact $t(Y)$ such that each theoretical fact s belongs to $t(Y)$, i.e., $s \in t(Y)$ if and only if all empirical correlates z of s belong to Y . In other

words, $t(Y)$ is the most comprehensive theoretical fact for which Y provides a complete empirical realization.

We hasten to add that this relational account of empirical theories as a relation $T \subseteq E \times S$ is seriously incomplete. Its essential flaw is that it does not allow us to distinguish between approximately true theories and false theories, i.e., theories that are completely off the mark. If a theory T is just a relation $T \subseteq E \times S$ relating symbolic and empirical facts, there is no room for asking if T is (approximately) correct or not. This is clearly not sufficient to model the way of how theories relate theoretical facts to often recalcitrant empirical facts. To overcome this shortcoming, it is expedient to rely once more on the insights encapsulated in Hertz's diagram. In other words, we propose to combine the insights of Hertz and Duhem to obtain a better model of scientific theorizing that comprises the advantages of both the Hertzian and the Duhemian accounts.

This is done as follows: Let us start over again from the domains PS and PE of theoretical facts and symbolic facts, respectively, endowed with maps $e: PS \xrightarrow{e} PE$ and $PE \xrightarrow{t} PS$ as before. That is to say, e and t are to be interpreted as Duhemian maps correlating symbolic facts and empirical facts as explained above. The new ingredient we are going to introduce in order to distinguish between (approximately) true theories and those that are plainly false is provided by the replacement of the trivial set theoretical order relation \subseteq_S on S and \subseteq_E on E by appropriate non-trivial order relations \leq_S and \leq_E on PS and PE , respectively, which reflect some theoretical or empirical intervention and processes as explained in our discussion of the Hertz diagram in section 2. More precisely this is explained in the following definition:

(5.4) *Definition.* (a) Assume $Y, Y^* \in PE$. Assume that there is an empirical process P or intervention such that the empirical fact Y is the initial state $P(i)$ of P , and Y^* is the final state $P(f)$ of P . It is further assumed that processes or interventions P, P', P'' can be concatenated associatively. Define $Y \leq Y^* :=$ there is a process P with initial state Y and final state Y^* .

(b) Assume $X, X^* \in PS$. Assume that there is a symbolic process P or intervention such that the symbolic fact X is the initial state $P(i)$ of P , and X^* is the final state $P(f)$ of P . It is further assumed that processes or interventions P, P', P'' can be concatenated associatively. Define $X \leq X^* :=$ there is a process P with initial state X and final state X^* .

The class of processes or interventions defined for symbolic and empirical facts render PS and PE order structures, to be denoted by (PS, \leq_S) and (PE, \leq_E) , respectively. From now on, PS and PE are assumed to be endowed with these interventional orders which differ from the set-theoretical orders \subseteq_S and \subseteq_E . In Hertz's terms, then, $X \leq X'$ is to read as ' X is a necessary consequent of X' ', and analogously $Y \leq Y'$ is to

be read as ‘ Y is a necessary consequent of X ’. Then the following Duhem-Hertz requirement makes sense:

(5.5) *Definition.* Let $T \subseteq E \times S$ be a relation of symbolical and empirical facts. Assume PE and PS endowed with interventional orders \leq_E and \leq_S respectively, $X \in PS$, $Y \in PE$. Let $PS \xrightarrow{e} PE$ and $PE \xrightarrow{t} PS$ defined by T . Then the theory T is said to satisfy the Duhem-Hertz condition iff for all $X \in PS$, $Y \in PE$ the following equivalence holds:

$$(5.6) \quad e(X) \leq_E Y \text{ IFF } X \leq_S t(Y).$$

In other words, the pair (t, e) is a Galois connection between (PS, \leq_S) and (PE, \leq_E) . More precisely, t is the upper (or right) adjoint, and e is the lower (or left) adjoint of this Galois connection.

Before we explain in some detail why theories satisfying (5.6) should be considered as (approximately) true let us note that instead of the set theoretical structures PS and PE it may be more expedient, more intuitive, and even less clumsy to replace PS and PE by ordered domains (U, \leq_U) and (V, \leq_V) . Then the Duhem-Hertz condition (5.6) simply requires that there are order-preserving maps $U \xrightarrow{e} V$ and $V \xrightarrow{t} U$ such that (t, e) defines a Galois connection between U and V in the sense of (5.2). This may be even further generalized by the assumption that U and V are categories in an adjoint situation (cf. Goldblatt 1978). That is to say, conceiving an empirical theory as an adjoint situation (F, G) between a category of symbolic facts U and a category V of empirical facts combines in a neat and natural way the classical insights of Hertz and Duhem.

As a summary of this section let us reformulate once more the basic thesis in somewhat different terms, assuming that an empirical theory is given as a Galois connection (t, e) between an ordered domain (U, \leq) of symbolic facts and an ordered domain (V, \leq) of empirical facts, i.e. the maps $U \xrightarrow{e} V$ and $V \xrightarrow{t} U$ satisfy the Galois equivalence

$$(5.7) \quad e(x) \leq a \quad \text{IFF} \quad x \leq t(a), \quad x \in U, a \in V.$$

Then we may conceive x as a *theoretical law* that may be considered as the blueprint for the building of a nomological machine or experimental apparatus $e(x)$ that produces the empirical fact a as its outcome. Then the Galois connection states:

The nomological machine $e(x)$ brings about the empirical fact a

IFF

The theoretical law x implies an idealized version $t(a)$ of a .

This yields another interpretation of the formal apparatus of Galois connection that renders plausible the claim why theories which satisfy the Galois connection should be considered as (approximately) true theories: such theories are approximately true since they ensure a relation between the empirical and the theoretical that captures the idea that an approximately true theory should approximately correspond to the facts.

6. Concluding Remarks

The leitmotif of this paper was the thesis that scientific theories are to be considered as *representations*, and, more generally, that the practice of science may be conceptualized as a representational practice. This idea is not new, and many have put forward it in many different ways. Philosopher-scientists such as Hertz and Duhem provide distinguished examples. Tapping some of their essential insights we hope to have rendered plausible the following theses: (i) representation is a complex concept in need of a theory, (ii) representations do not live in isolation. Rather, they may be *iterated* and *combined* in various ways, and (iii) representations do not ‘speak for themselves’. Rather, representations are in need of interpretation. A large part of scientific practice consists in interpreting and reinterpreting representations.

REFERENCES

- Duhem, P. (1906). *The Aim and Structure of Physical Theory*. Princeton: Princeton University Press, 1954.
- Gierz, G.; Hofmann, K.H.; Keimel, K.; Lawson, J.D.; Mislove, M.; Scott, D.S. (2003). *Continuous Lattices and Domains*. Cambridge: Cambridge University Press.
- Goldblatt, R. (1978). *Topoi – The Categorical Analysis of Logic*. Amsterdam and New York: North-Holland.
- Hacking, I. (1983). *Representing and Intervening*. Cambridge: Cambridge University Press.
- Hertz, H. (1894). *The Principles of Mechanics Presented in a New Form*. New York: Dover, 1956.
- Ibarra, A.; Mormann, T. (1997a). *Representaciones en la ciencia. De la invariancia estructural a la significatividad pragmática*. Barcelona: Ediciones del Bronce.
- (1997b). “Theories as Representations”, in A. Ibarra, T. Mormann (eds.), *Representations of Rationality*. Poznan Studies in the Philosophy of Science and the Humanities 61, Amsterdam and Atlanta, 65-92.
- (2000). “Una teoría combinatoria de la representación científica”, *Crítica*, vol. XXXII, N°. 95, 3-46.
- Latour, B. (1999). *Pandora’s Hope. Essays on the Reality of Science Studies*, Cambridge/Massachusetts: Harvard University Press.

- Lawvere, F. W.; Schanuel, S.H. (1996). *Conceptual Mathematics. A First Introduction to Categories*. Cambridge: Cambridge University Press.
- Laymon, R. (1982). "Scientific Realism and the Hierarchical Counterfactual Path from Data to Theory", in P. Asquith and T. Nickles (eds.), *PSA 1982*, vol. 1, Philosophy of Science Association, East Lansing, 107-121.
- Mac Lane, S. (1986). *Mathematics. Form and Structure*. New York: Springer.
- Melton, A.; Schmidt, D.A.; Strecker, G.E. (1985). *Galois Connections and Computer Science Applications*. Springer Lecture Notes in Computer Science 240. Berlin: Springer, 299-312.
- Strand, R.; Fjelland, R.; Flatmark, T. (1996). "In Vivo Interpretations of In Vitro Effect Studies with a Detailed Analysis of the Method of *in vitro* Transcription in Isolated Cell Nuclei", *Acta Biotheoretica* 44, 1-21.
- Strand, R. (1999). "Towards a Useful Philosophy of Biochemistry: Sketches and Examples", *Foundations of Chemistry* 1, 271-294.

Andoni IBARRA and **Thomas MORMANN** have worked together in a number of projects dealing with the role of representations in science, philosophy of science, and general epistemology. They published a book on this topic (*Representaciones en la Ciencia*, Barcelona, 1997) and served as editors of a number of anthologies on representational themes.

Andoni Ibarra moreover is presently coordinator of the Sanchez-Mazas Chair at the University of the Basque Country. Thomas Mormann holds a PhD in Mathematics and obtained his Habilitation in philosophy from the University of Munich (Germany).

ADDRESS: Dpt. of Logic and Philosophy of Science, University of the Basque Country, Av. Tolosa 70, 20018 Donostia-San Sebastián, Spain. E-mails: andoni.ibarra@ehu.es; ylxmomot@sf.ehu.es.

On the Analogy between Cognitive Representation and Truth

Mauricio SUÁREZ and Albert SOLÉ*

BIBLID [0495-4548 (2006) 21: 55; pp. 39-48]

ABSTRACT: In this paper we claim that the notion of cognitive representation (and scientific representation in particular) is irreducibly plural. By means of an analogy with the minimalist conception of truth, we show that this pluralism is compatible with a generally deflationary attitude towards representation. We then explore the extent and nature of representational pluralism by discussing the positive and negative analogies between the inferential conception of representation advocated by one of us and the minimalist conception of truth.

Key words: representation, inferential conception, truth, minimalism.

1. *The Inferential Conception of Scientific Representation*

In a scientific representation some source A —typically a model, a graph, an equation— is used to represent some target B —typically a system, entity or phenomenon. The inferential conception of scientific representation (Suárez 2004) rejects the view that scientific representation is a relation between A and B that answers solely to the properties of A and B . Instead representation in science is conceived as an intentional activity, which cannot be reduced to any objective relation between the objects that stand as sources and targets of the representation, and is best characterised by two ‘surface’ features: its intentional or representational force and its inferential capacity.

The inferential conception takes it that agents’ pragmatic purposes are essential in two different ways to the nature of the kind of cognitive representations one finds in science: i) as initial fixers of the representational *force* that points from A to B when A represents B , and ii) as defining the level of information, skill and competence required for an appropriate use of the representation, which in turn determines the *inferential capacities* of the source —i.e. it determines the inferences about B that can legitimately be carried out on the basis of reasoning about A . The representational force of a model within a practice, for instance, is typically initially fixed by stipulation and thereafter maintained by convention; the model’s inferential capacities are institutionally preserved through the practices of model-building in science. Both are essential ingredients of representation, though admittedly they do not cut deep: it is built into the notion of a cognitive representation that sources are representationally directed towards their targets; and similarly, that they allow or permit the carrying out of infer-

* We would like to thank two anonymous referees for helpful suggestions. Research towards this paper has been supported by projects HUM 2005-07187-C03-01, HUM 2005-04369, and PR 2004-0509 of the Spanish Ministry of Education and Science.



ences regarding these targets. These two features may then be considered among the most basic platitudes regarding cognitive or informative representation.¹

The inferential conception was explicitly linked from the start to ‘a deflationary or minimalist attitude and strategy towards the concept of scientific representation in analogy to deflationary or minimalist conceptions of truth’ (Suárez 2004, 770). We take it that a generally deflationary attitude towards any concept entails, roughly, an attempt at turning its platitudes into the defining conditions for the concept. It consequently involves the withdrawal of any further attempt at a ‘deeper’ or more substantive definition of the concept. This is precisely what the inferential conception invites us to do regarding the notion of representation as employed in science. Without prejudging the possibility of further platitudes, the inferential conception turns the two platitudes mentioned above into necessary conditions on representation, as follows (Suárez 2004, 773):

[inf]: A represents B only if (i) the representational force of A points towards B , and (ii) A allows competent and informed agents to draw specific inferences regarding B .

[inf] is not intended to capture one unique concept but a plurality of concepts of representation, since it leaves open a number of different possibilities for a completion into further necessary conditions or even necessary *and* sufficient conditions. Thus suppose that a further platitude x is found in the use of the notion of representation, which we might hope will complete the analytical definition of the concept. This can always be added to the set of necessary conditions established by [inf] in the following fashion:

[plural inf]: A represents B if and only if (i) the representational force of A points towards B , (ii) A allows competent and informed agents to draw specific inferences regarding B , and (iii) x .

First, note that the logical relationship between [inf] and [plural inf] is that the former is entailed, but does not entail, the latter. In proposing [plural inf] we are extending the original proposal into a more general schema which we characterise as *irreducibly* plural in the following sense. We assume that there are different members of [plural inf] with conditions x , y , z , etc, all of them legitimately defining a distinct concept of representation. Hence representation is a word that refers to several distinct concepts that share some but not all of its structure. This move turns the original proposal [inf] into a partial specification of the inferential conception of representation, which more precisely corresponds to the whole family of members of [plural inf]. We find it an advantage of the inferential conception that it allows for the possibility of maximal plurality—but it is clear to us that it then becomes imperative to try to locate

¹ Cognitive representations differ from mere stipulations or denotations, since they allow us to infer information about the intrinsic properties of targets that we would not be able to infer from any other arbitrary representation. Throughout we take scientific representation to be a subspecies of cognitive representation in general, and aim at characterising this most general notion. See Suárez 2004, Section 3, for a discussion.

the core conditions in virtue of which all these concepts fall under the same term. That is the job that the original proposal [inf] was designed to fulfil: identifying the two conditions that are conceptually necessary ('platitudinous') for any instance of scientific representation.

There is a particularly simple member of [plural inf] that achieves the logical closure of [inf] (and hence a full analytical definition of the concept):

[closed inf]: A represents B if and only if (i) the representational force of A points towards B and (ii) A allows competent and informed agents to draw specific inferences regarding B .

We are not suggesting to replace the original [inf] with [closed inf], since we consider [closed inf] only one case in a large family of interesting possibilities generated by [plural inf]. We are however, particularly keen to explore the features of this member, in particular with respect to the analogy with truth previously pointed out, since it might be considered the most *conservative* completion of the inferential conception and it restores the integrity of the concept of representation wholesale.²

Since we are thus providing a full analytical definition of the concept, it can now be questioned to what extent [closed inf] constitutes a deflationary or minimalist conception of representation. The main purpose of this paper is to defend the claim that it does. In other words we aim to show that even the simplest, or more conservative, member of [plural inf] satisfies the desiderata that we originally set for a generally deflationary attitude to representation.

We will defend this claim by developing the analogy with Crispin Wright's minimalist conception of truth. This should not be taken to entail support for any specific conception of truth, whether substantive or deflationary; in particular, it should not be taken as endorsement of Wright's 'inflationary' argument against Horwich's position (Wright 1992, Chapter 1; 2003, 337ff.). We do not need in this paper to defend any position on truth, since the issue is tangential to the discussion about cognitive representation.³

2. Minimalism and Deflationism about Truth

According to Wright, deflationism and minimalism share the contention that the conceptual analysis of truth must make essential reference to a couple of basic platitudes

² In other words, in this paper we make our commitment explicit to [plural inf], as a general schema of the inferential conception of representation. But we are not thereby committing ourselves to this particular version in lieu of the general schema [plural inf]. What we advocate instead is more philosophical research into various possible completions of this general schema and their relative virtues, as well as their domains of application.

³ For arguments sharply distinguishing truth and representation see Giere 1999 and forthcoming; Suárez 2003; Bailer-Jones 2003.

or a priori principles, ‘a suitably generalised form of one (or both) of the following two schemata’ (Wright 2003, 332):⁴

Equivalence Schema for propositions (ES): *It is true that P iff P*

Disquotational Scheme for sentences (DS): *‘P’ is true iff P*

The deflationist argues further that there is nothing substantial left to say about truth, both in its conceptual analysis and in its application: the traditional metaphysical disputes about the nature of truth (in particular whether truth is reducible to correspondence, coherence, justification, etc), are in fact about nothing substantial.

The minimalist view, by contrast, while accepting the essential role that (ES) and (DS) play in the analysis of the abstract concept of truth, takes it that there are further properties of propositions, or sentences, that realise or instantiate concretely this concept of truth. However, according to the minimalist these properties are not the same in all cases, and might well vary from domain to domain of discourse. Hence minimalism combines the advantages of deflationism with respect to the concept of truth with those of pluralism regarding its application in practice: ‘Minimalism thus incorporates a potential pluralism about truth, in the specific sense that what property serves as truth may vary from discourse to discourse’ (Wright 2003, 334).

It would be a mistake however to identify the crucial distinction between minimalism and deflationism with an exclusive emphasis upon the abstract *concept* as opposed to the concrete *property* of truth. In other words it would be wrong to characterise this disagreement as one about whether truth is an abstract concept. The minimalist accepts that in *every* concrete instantiation or realisation of the truth concept some further properties will obtain; in other words, that there is no pure, or unmediated, application of the abstract concept. Conversely the deflationist can argue that in every application of the concept, an additional property is instantiated, namely the trivial property of *falling under* the truth concept.⁵ Hence both can accept that every legitimate ascription of the truth predicate involves both the application of an abstract truth concept and its instantiation via a concrete property of propositions, or sentences.

Instead the essential distinction between these two positions is at the point of application: for the deflationist the property that instantiates the truth concept is the same in each and every application of the concept —and in no case a substantive property. For the minimalist by contrast, the properties instantiating truth are many and diverse, co-varying with the different domains of discourse. In terminology suited to this paper we could say that according to the minimalist truth has many ‘means’ of

⁴ In this paper we follow Wright’s terminology —though it is not universally accepted.

⁵ See for instance Horwich 1998, 141-4. Horwich distinguishes different sorts of properties, and argues that any one justifiably applied to the truth concept will likely not be substantive —so deflationism is not in danger.

application, while according to the deflationist it only ever has one ‘means’ —i.e. falling under the extension of the truth-concept.

3. *The Analogy between Representation and Truth*

Let us now return to the completion of the inferential conception ([closed inf]) that we are proposing to explore in this paper. First of all it must be stressed that the point of departure of this conception and the minimalist or deflationist theories of truth is exactly the same: taking some platitudes as the conditions which constitute the analytic definition of the abstract concept the theory is dealing with.

We begin by emphasising the distinction between the *means* and the *constituents* of representation (see Suárez 2003 for this distinction, as well as the arguments mentioned below). The constituents of representation are the necessary and sufficient conditions that define the concept, while the means of representation are the relations between A and B (i.e.: relational properties of A and B), actually employed by scientists in order to infer consequences about B on the basis of reasoning about A . Among the most widely used means are similarity (understood as the sharing of properties between A and B), and isomorphism (between the structures exemplified by A and B). The latter is typical in the domain of the mathematical sciences, while the former accounts for cases of iconic representation in the less quantitative sciences. An important part of the background to the inferential conception is the set of arguments that show that none of the typical representational means, such as similarity and isomorphism, are in fact constituents of representation. This consequently shows that the platitudes about representation (‘representational force’ and ‘inferential capacities’) cannot be reduced to any of the typical representational means, thus vindicating the deflationist attitude that motivates the inferential conception.

Nevertheless a weaker relation *does* hold, namely: all representational means (such as isomorphism and similarity) are concrete instantiations, or realisations, of one of the basic platitudes that constitute representation, namely ii) inferential capacities. For suppose that similarity obtains between A and B . Then A and B have some properties $\{a_1, a_2, \dots, a_n\}$ in common. It follows that anyone sufficiently competent and informed about the representation of B by A can infer on the basis of A that $\{a_1, a_2, \dots, a_n\}$ are instantiated in B . Similarly, suppose that A and B exemplify isomorphic structures $A' = \langle D, P_j^n \rangle$ and $B' = \langle E, T_j^n \rangle$; where D, E are the domains of objects in each structure and P_j^n and T_j^n are the n -place relations defined in the structure. A' and B' are isomorphic if and only if there is a one-to-one and onto mapping $f: D \rightarrow E$, such that for any n -tuple $(x_1, \dots, x_n) \in D: P_j^n[x_1, \dots, x_n]$ if and only if $T_j^n[f(x_1), \dots, f(x_n)]$. It follows that a competent agent informed about the isomorphism can in principle infer that B' possesses $T_j^n[f(x_1), \dots, f(x_n)]$ from the observation that A' possesses $P_j^n[x_1, \dots, x_n]$.

Hence every obtaining of similarity or isomorphism between the source and the target of a representation is ipso facto an instantiation of part of the abstract concept

of representation. And since similarity and isomorphism are distinct relations between \mathcal{A} and B , appropriate as means of representation in different domains, it follows that the realisations of scientific representation, like those of truth according to minimalism, are irreducibly plural.

But it remains to be seen if, along the lines of the analogy with minimalism, it is also the case that the constituents of representation can only be applied via some more concrete means —such as isomorphism and similarity. In other words we may ask if it is possible that [closed inf] applies on its own, *without* simultaneously instantiating some further property, or properties, of \mathcal{A} and B . Note how very doubtful this sounds. Since part (ii) of [closed inf] states that it must be possible for a competent and informed agent to infer some conclusions regarding B (i.e. its properties), on the basis of a consideration of \mathcal{A} 's properties, there must be some operative rule of inference (whether or not actively employed by the particular agent) between \mathcal{A} 's and B 's properties, but such rule would precisely qualify as a concrete means of representation. Hence, according to [closed inf] there can be no application of representation without the simultaneous instantiation of a particular set of properties of \mathcal{A} and B , and their relation.

A similar argument applies to part (i) of [closed inf]. This part states that the representational force of the source must point towards the target of the representation; it must be noted that [closed inf] leaves open the question of which particular conditions must be met in every concrete instance of representation in order to fix and preserve the representational force of the source. This allows these conditions to vary from one domain of representation to another. So, part (i) of [closed inf] expresses a generic condition which can only be instantiated if further conditions are met —and these further conditions will vary from context to context. But here the analogy with the minimalist theory is even stronger since a different analysis of the representational force will apply to each whole domain of representation (scientific, artistic, etc) in the same way as according to Wright every property that instantiates the abstract concept of truth applies throughout a whole domain of discourse.

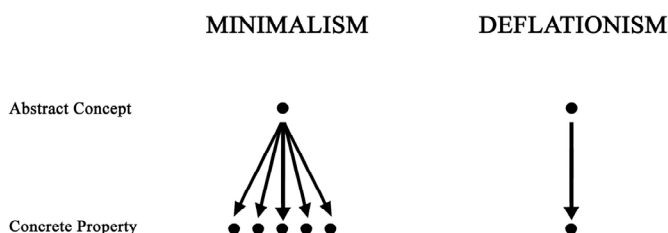
This is the core of the positive analogy between the inferential conception of representation in the simplest, or most conservative version [closed inf], and the minimalist conception of truth. In both cases an abstract concept is applied via a variety of concrete properties of the objects characteristically falling under the concept (propositions or sentences in the case of truth, sources and targets in the case of representation), but *which* concrete properties co-varies with the domains of application. The view similarly combines deflationism regarding the abstract concept with pluralism regarding the concrete property that serves to instantiate it.

We would like to point out in addition that the analogy brings out a particularly welcome feature of the inferential conception in the context of the present-day debates on the nature of scientific representation. Wright employs minimalism to explain away the metaphysical debates concerning different theories of truth (as correspondence, coherence, justification, etc). On the minimalist view these theories no longer characterise the concept of truth but its properties instead. It is then possible to show

that the different theories correspond to different sets of truth-instantiating properties in different domains, so every theory can feel vindicated in its own domain of discourse (and no theory is vindicated as a universal account of the concept of truth). Similarly the inferential conception explains away the quarrels between different theories of representation; for each theory now describes a different means of representation, appropriate in different domains (i.e. isomorphism being appropriate in the case of the most mathematical dynamical descriptions of nature, similarity appropriate for the less quantitative sciences). So these theories aim to characterise the means of representation in particular domains of scientific modelling; they do not characterise — nor should they be understood as trying to characterise— the constituents of representation. A long-standing dispute in the field is thereby resolved.

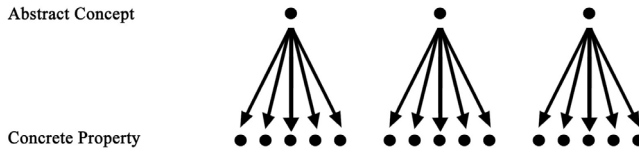
4. *The Limits of the Analogy*

The analogy between representation and minimalist truth is a good heuristic tool to explore the properties of the inferential conception, but it is not a perfect analogy. In this final section we point out two sources of negative analogy between the inferential conception and minimalism about truth. One of them lies precisely at the heart of what minimalism and deflationism share in common. Although minimalism and deflationism disagree about the plurality of ‘means’ of application, and specifically about whether there is only one or many, the deflationist and the minimalist agree on the other hand that there is only one concept of truth. What is debated is how to characterise this concept, and whether it has one or multiple realisations in terms of distinct properties, but the fact that there is only one such concept is not under discussion. The differences might be captured by means of the following diagram:



Hence minimalism allows for a plurality of concrete properties, but sticks to one abstract concept. Here the analogy with the inferential conception in general breaks down since [plural inf] allows (but not entails) that there might be a plurality of abstract concepts of representation: [closed inf] is just one of them. In other words the inferential conception of representation adheres to the following diagram:

INFERENCE CONCEPTION OF REPRESENTATION



The inferential conception is in this regard better conceived as a research programme for the development of alternative, and potentially competitive, notions of representation. In each domain of scientific discourse there might be different platitudes that need to be added to define the concept of representation appropriate for that domain. This patently finds no analogue in the minimalist conception of truth.

A second negative analogy we would like to mention between minimalist truth and the inferential conception concerns the possibility of turning these concepts into second-order properties. Such a move has already been developed with respect to minimalism. Michael Lynch (2001) for instance has developed what he calls alethic functionalism; essentially this is the claim that truth is a second order property of propositions, or sentences, namely, the property of having a first-order property that plays the truth-role. This accommodates the minimalist insight nicely, since it allows us to account for the plurality of concrete ‘truth’ properties in different domains: coherence for juridical science, for instance, correspondence for ordinary factual discourse about macroscopic objects perhaps, etc. What these properties share according to the functionalist is not so much their instantiating the same abstract concept of truth, but their playing the truth-role in their domain (a role characterised at the very least by the fulfilment of the platitudes). The generic property of playing the truth-role is thus realised by different properties of propositions, or sentences, in different domains; however, according to the functionalist, truth is precisely this 2nd order property.

The move to a 2nd order property theory of truth is of course controversial and in fact changes considerably the nature of the minimalist project. It is not obvious that an analogous functionalist theory for representation would satisfy our conditions for a deflationary attitude. But the analogy with the inferential conception breaks down at this stage anyway, since the latter cannot be considered a 2nd order property of source-target pairs. For consider what this would entail —roughly that there are particular properties of chosen source-target pairs that play the representation-role in each domain, and that it is the generic 2nd order property of playing that role (a role characterised at least by the platitudes about representation) that constitutes ‘representation’. The inferential capacities of the source (part (ii) of [closed inf]) could indeed be taken to describe a representation-role across each of the domains (fulfilled by a different means of representation in each domain as we have seen). However the problem is

that part (i) states that the representational force must flow from the source towards the target, and this is not per se a property of the objects that function as source or target, nor is it therefore a relation between them. So there is no room here to exploit the idea of a 2nd order property of 1st order properties of the objects related, and the functionalist theory seems to lack an analogue in the inferential conception.

5. Conclusions

We have explored the positive and negative analogy between the inferential conception of cognitive representation (which we propose to formally represent by [plural inf] in general) and the minimalist conception of truth defended by Crispin Wright. The analogy is introduced for heuristic purposes, in order to display the kind of pluralism that we take cognitive representation to possess. We do not wish to establish any deeper theoretical link between representation and truth, and we have already declared our neutrality regarding the nature of truth.

The point of the positive analogy is to strengthen and to clarify the distinction between the means and the constituents of representation. The latter are given by the platitudes of representation and define the abstract concept, while the former correspond to the set of concrete properties (such as isomorphism, similarity, homology, etc) that instantiate the abstract concept —each property being the characteristic form of instantiation in its corresponding domain. While the point of the negative analogy is, first, to show that the inferential conception admits a plurality (in fact a whole family) of abstract concepts of representation —in contrast to minimalism which takes truth to be a univocal concept. Second, the attempts to reformulate minimalism as a functionalist 2nd order property of truth would seem to lack any possible analogue in the case of the inferential conception of representation.

To conclude we see this analogy as heuristic reinforcement of the view that the abstract constituents of representation are instantiated, or realised, partly through the concrete means of representation. In turn this vindicates the claim that the kinds of cognitive representation characteristic of science are abstract relations that obtain in practice through isomorphism, similarity, homology, and so on —but which should not be identified with any of them.

REFERENCES

- Bailer-Jones, D. M. (2003). “When Scientific Models Represent”, *International Studies in the Philosophy of Science* 17: 59-74.
- Giere, Ronald (1999). *Science Without Truth*. University of Chicago Press.
- (forthcoming). *Scientific Perspectives*. University of Chicago Press.
- Horwich, Paul (1998). *Truth*. Oxford: Blackwell.
- Lynch, Michael (2001). “A Functionalist Theory of Truth”, in Michael Lynch (ed.), *The Nature of Truth*. Cambridge, MA: MIT Press.
- Suárez, Mauricio (2003). “Scientific Representation: Against Similarity and Isomorphism”, *International Studies in the Philosophy of Science* 17 (3): 225-244.
- (2004). “An Inferential Conception of Scientific Representation”, *Philosophy of Science* 71: 767-779.

Wright, Crispin (1992). *Truth and Objectivity*. Cambridge, MA: Harvard University Press.
——— (2003). *Saving the Differences*. Cambridge, MA: Harvard University Press.

Mauricio SUÁREZ is Associate Professor (Profesor Titular) in Logic and Philosophy of Science at Complutense University of Madrid. He previously taught at Oxford, St. Andrews and Bristol Universities, and was postdoctoral research fellow at Northwestern University. He holds a BSc in Astrophysics (Edinburgh) and PhD in Philosophy of Science (LSE). He has published widely on models and representation in science, on dispositions and causality in quantum physics, and on general scientific epistemology.

ADDRESS: Department of Logic and Philosophy of Science, Faculty of Philosophy, Complutense University, 28040 Madrid, Spain. E-mail: msuarez@filos.ucm.es.

Albert SOLÉ graduated in both Physics and Philosophy from the University of Barcelona. He is currently doing his PhD research work in Philosophy of Science at Complutense University of Madrid. His interests range from issues in general methodology of science to more particular topics in the philosophy of physics, particularly in the foundations of quantum mechanics and Bohm's theory.

ADDRESS: Department of Logic and Philosophy of Science, Faculty of Philosophy, Complutense University, 28040 Madrid, Spain. E-mail: asole@filos.ucm.es.

Scientific Representation and the Semantic View of Theories

Roman FRIGG

BIBLID [0495-4548 (2006) 21: 55; pp. 49-65]

ABSTRACT: It is now part and parcel of the official philosophical wisdom that models are essential to the acquisition and organisation of scientific knowledge. It is also generally accepted that most models represent their target systems in one way or another. But what does it mean for a model to represent its target system? I begin by introducing three conundrums that a theory of scientific representation has to come to terms with and then address the question of whether the semantic view of theories, which is the currently most widely accepted account of theories and models, provides us with adequate answers to these questions. After having argued in some detail that it does not, I conclude by pointing out in what direction a tenable account of scientific representation might be sought.

Keywords: Scientific representation, models, semantic view of theories, isomorphism, similarity.

1. Introduction

Models are of central importance in many scientific contexts, where they play an essential role in the acquisition and organisation of scientific knowledge. We often study a model to discover features of the thing it stands for. For instance, we study the nature of the hydrogen atom, the dynamics of populations, or the behaviour of polymers by studying their respective models. But for this to be possible models must be representational. A model can instruct us about the nature of reality only if we assume that it represents the selected part or aspect of the world that we investigate.¹ So if we want to understand how we learn from models, we have to come to terms with the question of how they represent.

Although many philosophers, realists and antirealists alike, agree with a characterisation of science as an activity aiming at representing parts of the world,² the issue of scientific representation has not attracted much attention in analytical philosophy of science until recently. So the first step towards a satisfactory account of scientific representation is to be clear on the questions that such an account is supposed to deal with and on what would count as satisfactory answers. I address this issue in the next section. In the remainder of the paper I discuss currently available accounts of theories and models and argue that, whatever their merits on other counts, they do not provide us with satisfactory answers to the problems a theory of representation has to solve.

¹ This is not to say that models are ‘mirror images’ or ‘transcripts’ of nature. Representing need not (and usually does not) amount to copying.

² Cartwright (1999, esp. Ch. 8), Giere (1988; 1999; 2004), Hughes (1997), Kitcher (1993), Morgan and Morrison (1999), Psillos (1999), Redhead (2001), Suppe (1989), van Fraassen (1980; 1997; 2002), to mention just a few.



2. *The Three Conundrums of Scientific Representation*

A theory of scientific representation has to come to terms with (at least) three conundrums. The first one is the ontology of models: what kinds of objects are models? Are they structures in the sense of set theory, fictional entities, concrete objects, descriptions, equations or yet something else? I refer to this issue as the ‘*ontological puzzle*’.

The second and the third conundrum are concerned with the semantics of models. Models are representations of a selected part or aspect of the world (henceforth ‘target system’). But in virtue of what is a model a representation of something else?³ To appreciate the thrust of the question it is helpful to consider the analogous problem with pictorial representation, which Flint Schier eloquently dubbed the ‘enigma of depiction’ (1986, 1). When seeing, say, Pissarro’s *Boulevard des Italiens* we immediately realise that it depicts one of the glamorous streets of *fin de siècle* Paris. Why is this? The symbolist painter Maurice Denis famously took wicked pleasure in reminding his fellow artists that a painting, before being a nude or a landscape, essentially is a flat surface covered with paint, a welter of lines, dots, curves, shapes, and colours. The puzzle then is this: how do lines and dots represent something outside the picture frame? Slightly altering Schier’s congenial phrase, I refer to the problem of how models represent their targets as the ‘*enigma of representation*’ (‘enigma’, for short).

The third conundrum is what I call the ‘*problem of style*’, which comes in a factual and a normative variant. Not all representations are of the same kind. In painting this is so obvious that it hardly deserves mention. An ink drawing, a wood cut, a pointillist painting, or a geometrical abstraction can represent the same scene in very different ways. This pluralism is not a prerogative of the fine arts. The representations used in the sciences are not all of the same kind either. Bill Phillips’ hydraulic machine and Hicks’ mathematical models both represent a Keynesian economy but they use very different devices to do so; and Weizsäcker’s liquid drop model represents the nucleus of an atom in a manner that is very different from the one in the shell model. As in painting, there seems to be a variety of representational styles in science. But what are these styles (or ‘modes of representation’)? A theory of representation has to come up with a taxonomy of different styles and provide us with a characterisation of each of them. This is the factual aspect of the problem of style.

A further aspect of the problem of style is the normative question of whether there is a distinction between scientifically acceptable and unacceptable styles. One might be willing to grant that there are different representational strategies but still hold that only some of them truly deserve the label ‘scientific’. Are there any constraints on the choice of styles of representation in science?

In sum, a theory of representation has to come to terms with three conundrums, two semantic, and one ontological. I do not claim that this list is exhaustive; but I

³ One could render this question more precise by asking ‘what fills the blank in “*M* is a scientific representation of *T* iff ___”, where “*M*” stands for “model” and “*T*” for “target system”?’ However, it is not obvious that there are necessary and sufficient conditions to be had here and it does not seem appropriate to regard an account of representation as successful only if it provides such conditions.

think that whatever list of questions one might put on the agenda of a theory of scientific representation, these three will be among them and they will occupy centre stage in the discussion.

Many answers to these questions are in principle possible and it is far from clear what would count as an acceptable theory of scientific representation. But there are (at least) two requirements that any such theory should satisfy.

First, learning from models. Scientific models represent things in a way that allows us to acquire knowledge about them. We study a model and thereby discover features of the thing it stands for. Every acceptable theory of scientific representation has to account for this interplay between knowing and representing.⁴

Second, the possibility of misrepresentation. A tenable theory of scientific representation has to be able to explain how misrepresentation is possible.⁵ Misrepresentation is common in science. Some cases of misrepresentation are, for all we know, plain mistakes (e.g. ether models). But not all misrepresentations involve error. Many models are based on idealising assumptions of which we know that they are false. Nevertheless these models are representations. A theory that makes the phenomenon of misrepresentation mysterious or impossible must be inadequate.

Where do we stand on these issues? Over the last four decades the semantic view of theories has become the orthodox view on models and theories. Although it has not explicitly been put forward as an account of scientific representation, representation-talk is ubiquitous in the literature on the semantic view and its central contentions clearly bear on the issue. So it seems to be a natural starting point to ask whether the semantic view provides us with adequate answers to the above questions. I argue that it does not. Whatever the semantic view may have to offer with regards to other issues, it does not serve as a theory of scientific representation.

3. *The Structuralist Conception of Models*

There are two versions of the semantic view of theories, one based on the notion of structural isomorphism and one based on similarity. I will now focus on the former and return to the latter in section 8.

At the core of the first version of the semantic view lies the notion that models are structures. A structure $\mathcal{S} = \langle U, O, R \rangle$ is a composite entity consisting of (i) a non-empty set U of individuals called the domain (or universe) of the structure \mathcal{S} , (ii) an indexed set (i.e. an ordered list) O of operations on U (which may be empty), and (iii) a non-empty indexed set R of relations on U . In what follows I will omit opera-

⁴ This is in line with Morgan and Morrison who regard models as ‘investigative tools’ (1999, 11) and Swoyer who argues that they have to allow for what he calls ‘surrogate reasoning’ (1991, 449).

⁵ This condition is adapted from Stich and Warfield (1994, 6-7), who suggest that a theory of mental representation should be able to account for misrepresentation.

tions and take structures to be a domain endowed with certain relations. This is can be done without loss of generality because operations reduce to relations.⁶

For what follows it is important to be clear on what we mean by ‘individual’ and ‘relation’ in this context. To define the domain of a structure it does not matter what the individuals are—they may be whatever. The only thing that matters from a structural point of view is that there are so and so many of them.⁷ Or to put it another way, all we need is dummies or placeholders.

Relations are understood in a similarly ‘deflationary’ way. It is not important what the relation ‘in itself’ is; all that matters is between which objects it holds. For this reason, a relation is specified purely extensionally, that is, as class of ordered n -tuples and the relation is assumed to be nothing over and above this class of ordered tuples. Thus understood, relations have no properties other than those that derive from this extensional characterisation, such as transitivity, reflexivity, symmetry, etc.

This leaves us with a notion of structure containing dummy-objects between which purely extensionally defined relations hold.⁸

The crucial move is to postulate that scientific models are structures in exactly this sense. In this vein Suppes declares that ‘the meaning of the concept of model is the same in mathematics and the empirical sciences’ (1960a, 12). Van Fraassen posits that a ‘scientific theory gives us a family of models to represent the phenomena’, that ‘[t]hese models are mathematical entities, so all they have is structure [...]’ (1997, 528-99) and that therefore ‘[s]cience is [...] interpreted as saying that the entities stand in relations which are transitive, reflexive, etc. but as giving no further clue as to what those relations are’ (1997, 516). Redhead claims that ‘it is this abstract structure associated with physical reality that science aims, and to some extent succeeds, to uncover [...]’ (2001, 75). And French and Ladyman affirm that ‘the specific material of the models is irrelevant; rather it is the structural representation [...] which is important’ (1999, 109).⁹

⁶ See Boolos and Jeffrey (1989, 98-99). Basically the point is that an operation taking n arguments is equivalent to a $n + 1$ place relation.

⁷ This point is clearly stated in Russell (1919, 60).

⁸ There is a controversy over whether these structures are Platonic entities, equivalence classes, or modal constructs. For what follows it does not matter what stance one takes on this issue. See Dummett (1991, 295ff.), Hellman (1989), Redhead (2001), Resnik (1997), and Shapiro (2000, Ch. 10) for different views on this issue.

⁹ Further explicit statements of this view include: Da Costa and French (1990, 249), Suppes (1960b, 24; 1970, Ch. 2 pp. 6, 9, 13, 29), and van Fraassen (1980, 43, 64; 1991, 483; 1995, 6; 1997, 516, 522; 2001, 32-3). This is not to deny that there are differences between different versions of the semantic view. The precise formulation of what these models are varies from author to author. A survey of the different positions can be found in Suppe (1989, 3-37). How these accounts differ from one another is an interesting issue, but for present purposes nothing hinges on it. As Da Costa and French (2000, 119) —correctly, I think— remark, ‘[i]t is important to recall that at the heart of this approach [i.e. the semantic approach as advocated by van Fraassen, Giere, Hughes, Lloyd, Thompson, and Suppe] lies the fundamental point that theories [construed as families of models] are to be regarded as *structures*’ (original emphasis).

In keeping faithful to the spirit of this take on models, proponents of the semantic view posit that the relation between a model and its target system is isomorphism. As I mentioned at the beginning, the semantic view has not explicitly been put forward as a theory of representation.¹⁰ But given the general outlook of this approach, one might plausibly attribute to it the following account of representation:

(SM) A scientific model S is a structure and it represents the target system T iff T is structurally isomorphic to S .¹¹

I refer to this as the *structuralist view of models*. This view comes in grades of refinement and sophistication. What I have presented so far is its simplest form. The leading idea behind its ramifications is to replace isomorphisms by less restrictive mappings such as embeddings, partial isomorphisms, or homomorphisms. This undoubtedly has many technical advantages, but it does not lessen any of the serious difficulties that attach to (SM). For this reason, I consider the structuralist view in its simplest form throughout and confine my discussion of these ramifications to section 8, where I spell out how the various shortcomings of (SM) surface in the different ramified versions.

The question we have to address is whether (SM) provides us with a satisfactory answer to the three conundrums of scientific representation. The bulk of my discussion will be concerned with the enigma (sections 4 to 6) and I argue that (SM) is inadequate as a response to this problem. In section 7 I discuss the problem of style and conclude that (SM) fares only marginally better when understood as an answer to this problem. And what about the ontological claim? Are models structures? As I point out in section 9, it is a by-product of the discussion in sections 4-6 that this is not tenable either. Models involve, but are not reducible to structures.

4. *Structuralism and the Enigma I: Isomorphism Is Not Representation*

The arguments against (SM) as an answer to the enigma fall into two groups. Criticisms belonging to the first group, which I will be dealing with in this section, aim to show that scientific representation cannot be explained in terms of isomorphism. Arguments belonging to the second group regard the very notion of there being an isomorphism between model and target as problematic and conclude that in order to make sense of isomorphism claims structuralists have to tack on elements to their account of representation that they did not hitherto allow for. I discuss these objections in sections 5 and 6.

¹⁰ Recently, van Fraassen (2004) and French (2003) have paid some attention to the issue of representation from the perspective of semantic view of theories. However, no systematic account of representation emerges from their discussions.

¹¹ This view is extrapolated from van Fraassen (1980, Ch. 3; 1989, Ch. 9; 1997), French and Ladyman (1999), Da Costa and French (1990), French (2000), and Bueno (1997 and 1999), among others. Van Fraassen, however, adds pragmatic requirements —I shall come to these below.

The first and simple reason why representation cannot be explained in terms of isomorphism is that the latter has the wrong formal properties: isomorphism is symmetric and reflexive while representation is not.¹²

Furthermore, structural isomorphism is not sufficient for representation because in many cases neither one of a pair of isomorphic objects represents the other. Two copies of the same photograph, for instance, are isomorphic to one another but neither is a representation of the other.¹³ A corollary of this is that (SM) is unable to correctly fix the extension of a representation. It is a matter of fact that the same structure can be instantiated in different systems. For instance, a pendulum and certain kinds of electric circuits instantiate the same structure (Kroes 1989). In cases like this the model of the pendulum is isomorphic both to the pendulum and to the circuit. But it only represents the pendulum and not the circuit. Hence, isomorphism is too inclusive a concept to account for representation.

These criticisms suggest that (SM) is overly 'purist' in stipulating that representation has to be accounted for *solely* in terms of isomorphism, as all these problems vanish when we include intentional users in the definition of representation:¹⁴

(SM') The structure S represents the target system T iff T is structurally isomorphic to S and S is intended by a user to represent T .

This appears to be a successful move since (SM') is not vulnerable to the above objections. However, the move is so straightforward that it should make us suspicious. I agree that users do play an essential role in scientific representation; but merely tacking on intentions as a further condition is question begging. To say S is turned into a representation because a scientist intends S to represent T is a paraphrase of the problem rather than a solution. Consider an analogous problem in the philosophy of language: by virtue of what do some words refer? Merely saying that speakers intend words to refer to this or that is not an answer. Of course they do. What we really want to know is what is involved in a speaker establishing reference and a good deal of philosophy of language is an attempt to come to terms with this question. So what we have to understand is how a scientist comes to use S as a representation of T and to this end much more is needed than a blunt appeal to intentions.

Moreover, when we look at how (SM') solves the above-mentioned problems we realise that isomorphism has become irrelevant in explaining why S represents T as it is the appeal to intention that does all the work. Rather, isomorphism regulates the way in which the model has to relate to its target. Such regimentation is needed because an account of representation solely based on intention allows that everything

¹² This argument has been levelled against the similarity theory of pictorial representation by Goodman (1968, 4-5) and has recently been put forward against the isomorphism view by Suárez (2003).

¹³ This problem cannot be solved by requiring that models are structures and that targets are objects in the world because some models represent other models just as some pictures represent other pictures.

¹⁴ This is explicitly held by van Fraassen (1994, 170; 1997, 523 and 525).

can represent just about everything else by a mere act of *fiat*, which cannot be right as on such account we cannot explain how we learn from a model about the target.

However, when used in this way, isomorphism is put forward as an answer to the problem of style rather than the enigma: it imposes constraints on what kinds of representations are admissible but it does not contribute to explaining where a model's representational power comes from. Whether isomorphism is a sensible constraint to impose on the way in which a model represents will be discussed in section 7.

5. Structuralism and the Engima II: The Abstractness of Structural Claims

Isomorphism is a relation that holds between two structures and not between a structure and a piece of the real world *per se*. Hence, if we are to make sense of the claim that model and target are isomorphic we have to assume that the target exhibits a structure. What is involved in this assumption? Using a particular notion of abstraction I argue that structural claims do not 'stand on their own' in that a structure *S* can represent a system *T* only with respect to a certain description. As a consequence, descriptions cannot be omitted from an analysis of scientific representation and one has to recognise that scientific representation cannot be explained solely in terms of structures and isomorphism.

Some concepts are more abstract than others. *Playing a game* is more abstract than *playing chess* or *playing soccer* and *travelling* is more abstract than *sitting in the train* or *riding a bicycle*. What is it for one concept to be more abstract than another? Cartwright (1999, 39) provides us with two conditions:

First, a concept that is abstract relative to another more concrete set of descriptions never applies unless one of the more concrete descriptions also applies. These are the descriptions that can be used to "fit out" the abstract description on any given occasion. Second, satisfying the associated concrete description that applies on a particular occasion is what satisfying the abstract description consists in on that occasion.

Consider the example of *travelling*. The first condition says that unless I either sit in the train, drive a car, or pursue some other activity that brings me from one place to another I am not travelling. The second condition says that my sitting in a train right now *is* what my travelling consists in.

I now argue that *possessing a structure* is abstract in exactly this sense and it therefore does not apply without some more concrete concepts applying as well.

What is needed for something to have a certain structure *S* is that it consists of a set of individuals and that these enter into certain relations. Trivially, this implies that for it to be the case that *possessing a structure* applies to a system, *being an individual* must apply to some of its parts and *standing in a relation* to some of these. The crucial thing to realise at this point is that *being an individual* and *being in a relation* are abstract on the model of *playing a game*.

The applicability of *being an individual* depends on whether other concepts apply as well. What these concepts are depends on the context and the kinds of things we are dealing with (physical objects, persons, social units, etc). This does not matter; the salient point is that whatever the circumstances, there are *some* notions that have to apply

in order for something to be an individual. As an example consider ordinary medium-size physical objects. A minimal condition for such a thing to be an individual is that it occupies a certain space-time region. For this to be the case it must have a surface with a shape that sets it off from its environment. This surface in turn is defined by properties such as impenetrability, visibility, having a certain texture, etc. If we change scale, other properties may become relevant; but in principle nothing changes: we need certain more concrete properties to obtain in order for something to be an individual. If something is neither visible nor possesses shape, mass, or charge, then it cannot be treated as an individual.

And similarly with *being in a relation*. For it to be the case that two objects enter into a relation it also has to be the case that, say, one is hotter than, greater than or more beautiful than the other. In other words, *being in a relation* only applies if either *being hotter than*, *being greater than*, or *being more beautiful than* applies as well. *Being hotter than* is what being in a relation consists in on a particular occasion. Therefore, *being in a relation* is abstract in the above sense.

From this I conclude that *possessing a structure* does not apply unless some more concrete description of the target system applies as well. Naturally, this dependence on more concrete descriptions carries over to isomorphism claims. If we claim that T is isomorphic to S then, trivially, we assume that T has a structure S_T , which enters into the isomorphism with S . This assumption, however, presupposes that there is a more concrete description that is true of the system.

Let me end this section with a remark about supervenience. It may seem that the use of abstract concepts is somewhat far-fetched and that the same point could be made in a more elegant way by appeal to supervenience: structures supervene on certain non-structural base properties and hence one cannot have structures without also having these base properties. Details aside, I think that this is a valid point as far as the argument of this section goes. However, in the next section I argue that structures are not unique in the sense that the same object can exhibit different non-isomorphic structures. This is incompatible with supervenience because supervenience requires that any change in the structural properties be accompanied by a suitable change in the base properties. Abstraction does not require such a tight connection between structures and the concrete properties on which they rest.

6. *Structuralism and the Enigma III: The Chimera of the One and Only Structure of Reality*

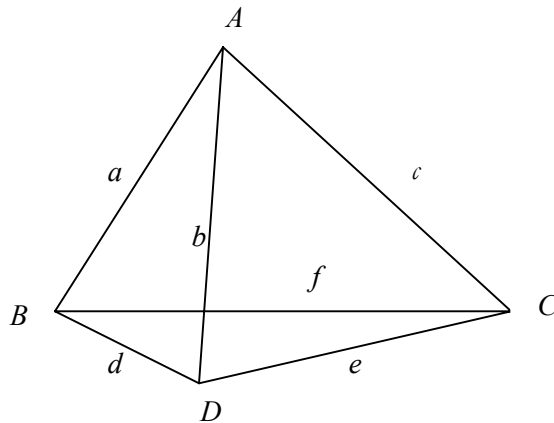
The main contention of this section is that a target system does not have a unique structure; depending on how we describe the system it exhibits different, non-isomorphic structures. If a system is to have a structure it has to be made up of individuals and relations. But the physical world does not come sliced up with the pieces having labels on their sleeves saying ‘this is an individual’ or ‘this is a relation’.¹⁵ What we recognise as individuals and what relations hold between these depends, in part at

¹⁵ And even if there is something like an ‘ultimate’ structure of reality, it is not this structure that most scientific models aim to represent.

least, on how we conceptualise the system. Because different conceptualisations may result in different structures there is no such thing as the one and only structure of a system. Needless to say, there are ways of ‘cutting up’ a system that seem simple and ‘natural’, while others may seem contrived. But what seems contrived from one angle may seem simple from another one and from the viewpoint of a theory of scientific representation any is as good as any other.^{16, 17}

My argument in support of this claim is inductive, as it were. In what follows I discuss examples from different contexts and show how the structure of the system depends on the description we choose. These examples are chosen such that the imposition of different structures only relies on very general features of the systems (e.g. their shape). For this reason, it is easy to carry over the strategies used to other cases. From this I conclude that there is at least a vast class of systems for which my claims bear out, and that is all I need.

The methane molecule (CH_4) consists of four hydrogen atoms forming a regular tetrahedron (see the figure below) and a carbon atom located in the middle. In many scientific contexts (e.g. collisions or the behaviour of a molecule *vis a vis* a semipermeable membrane) only the shape of the molecule is relevant. What is the structure of the shape of methane?



¹⁶ This position is compatible with, but does neither presuppose nor imply any form of metaphysical anti-realism or internal realism. I am only arguing for the much weaker claim that things do not have a unique structure.

¹⁷ This point, though pulling in the same direction, is not equivalent to Newman’s theorem, which, roughly, states that any set can be structured in any way one likes subject to cardinality constraints (Newman 1928, 144). This theorem is a formal result turning on the fact that relations are understood extensionally in set theory and that therefore a domain can be structured by putting objects into ordered n -tuples as one likes. What I argue is that a system can exhibit different *physically relevant* structures, i.e. structures that are not merely formal constructs but reflect the salient features of the system. I am aware of the fact that this is a somewhat vague characterisation and I rely on the subsequent examples to clarify the point.

To apply our notion of structure we need a set of individuals and relations.¹⁸ It seems a natural choice to regard the vertices as objects and take the edges to define relations. As a result we obtain the structure T_V which consists of a four-object domain $\{A, B, C, D\}$ and the relation L ($Lxy = 'x$ is connected to y by a line'), which has the extension $\{(A, B), (A, C), (A, D), (B, C), (B, D), (C, D)\}$.

However, this is neither the only possible nor the only natural choice. Why not consider the edges as the objects and the vertices as defining the relations? There is nothing in the nature of vertices that makes them more 'object-like' than edges. Following this idea we obtain the structure T_E with a domain consisting of the six edges $\{a, b, c, d, e, f\}$ and the relation I ($Ixy = 'x$ and y intersect'), which has the extension $\{(a, b), (a, c), (a, d), (a, f), (b, c), (b, d), (b, e), (c, e), (c, f), (d, f), (d, e), (e, f)\}$.

The upshot of this is that methane exemplifies a certain structure only with respect to a certain description and that there is no such thing as *the* structure of methane. And this is by no means a peculiarity of this example. The argument only relies on general geometrical features of the shape of methane and can easily be carried over to other objects.

Another straightforward example illustrating my claim is the solar system, which only has the structure that we usually attribute to it¹⁹ when we describe it as an entity consisting of ten perfectly spherical spinning tops with a spherical mass distribution. No doubt, this is a natural and in many respects useful way to describe this system, but it is by no means the only one. Why not consider the individual atoms in the system as basic entities? Or why not adopt a 'Polish' stance and also take the mereological sums of some planets as objects? There are many possibilities and each of these leads to a different structure.

The problem becomes even more pressing when we also take idealisations into account. As an example consider one of the earliest, and by now famous, ecological models. This model postulates that the growth of a population is given by the so-called logistic map: $x' = Rx(1 - x)$, where x is the population density in one generation and x' in the next; R is the growth rate. For this to be a representation, the structuralist has to claim that the structure S_L , which is defined by the logistic map, is isomorphic to the structure of the population under investigation. But this is only true when we describe this population in particular way. As Hofbauer and Sigmund (1998, 3) point out, in many ecosystems thousands of species interact in complex patterns depending on the effects of seasonal variations, age structure, spatial distribution and the like. Nothing of this is visible in the model. It is just the net effect of all interactions that is accounted for in the last term of the equation ($-Rx^2$). And a similarly radical move is needed when it comes to defining the objects of the structure. An obvious choice would be individual animals. But one readily realises that this would lead to intractable sets of equations. The 'smart' choice is to take generations rather than individual insects as objects. We furthermore have to assume that the generations are non-

¹⁸ This example is discussed in Rickart (1995, 23, 45).

¹⁹ For a detailed discussion of this structure see Balzer *et al.* (1987, 29-34, 103-8, 180-91).

overlapping, reproduce at a constant average rate (reflected in the magnitude of R) and in equidistant discrete time steps. Hence we have to describe the system in this particular way for it to exhibit the structure we are dealing with; and if we choose different descriptions involving different modelling assumptions we obtain different structures.

To end the discussion of the enigma, let me briefly mention a possible objection: all I have said so far is wrongheaded from beginning to end because it misconstrues the nature of the target system. I have assumed that what a model represents is an object (or event) of some sort. But, so the objection goes, this is mistaken. What a model ultimately represents is not an object, but a data model.²⁰ Space constraints prevent me from discussing this objection in detail, so let me just state that I think that this objection is wrong for the reason which Bogen and Woodward (1988) have pointed out: models represent phenomena, not data.²¹

7. *Structuralism and the Problem of Style*

So far I have argued that (SM) is untenable as a response to the enigma. Before drawing some general conclusions from this, I want to address the question of whether (SM) fares better as a response to the problem of style (this section) and argue that amended versions face, *mutatis mutandis*, the same difficulties (section 8).

The problem of style in its factual variant is concerned with modes of representation: what different ways of representing a target are there? For sure, isomorphism is one possible answer to this question; one way of representing a system is to come up with a model that is structurally isomorphic to it. This is an uncontroversial claim, but also not a very strong one.

The emphasis many structuralists place on isomorphism suggests that they do not regard it merely as one way among others to represent something. What they seem to have in mind is the stronger, normative contention that a representation *must* be of that sort.

This contention is mistaken. First, it is a common place that many representations are inaccurate in one way or another and as a consequence their structure is not isomorphic to the structure of their respective target systems. Second, it runs counter to the second condition of adequacy, namely that misrepresentation must be possible. To require that a model must be isomorphic to its target amounts to saying that only accurate representations count as representations and to ruling out cases of misrepresentation (i.e. cases in which isomorphism fails) as non-representational, which is unacceptable.

Structuralists may counter that this reading of the claim that representation involves isomorphism is too strong and argue that it is only something like a regulative ideal: as science progresses, its models have to become isomorphic to their target sys-

²⁰ See van Fraassen (1981, 667; 1985, 271; 1989, 229; 1997, 524; 2001, 31; 2002, 164, 252) and French (1999, 191-92).

²¹ See also Woodward (1989), McAllister (1997) and Teller (2001).

tems. This claim, however, falls outside the scope of a theory representation for it is just convergent realism in structuralist guise. But questions concerning realism or anti-realism should be kept apart from the issue of scientific representation. Convergent realism is a position one can hold, but as a view on representation it is besides the point. Representations can be realistic, but they do not have to be. Scientific modelling does not always amount to pointing a mirror towards things and making convergent realism a part of a theory of representation is neither necessary nor desirable.

8. *Why Other Accounts Do Not Fare Better*

The leading idea of amended versions of (SM) is to relax the isomorphism requirement and use a less restrictive mapping. Some prominent suggestions include embedding (Redhead 2001), homomorphism (Mundy 1986), and partial structures (French and Ladyman 1999).

Whatever advantages these mappings enjoy over isomorphism in other contexts, it is not difficult to see that none of them resolves any of the above-mentioned difficulties. In order to set up any of these mappings between the model and the target we have to assume that the target exemplifies a certain structure and therefore these views are also subject to the criticisms levelled against (SM) in sections 5 and 6. And with regards to the problems mentioned in section 4 amended versions fare only marginally better. None of these mappings is necessary for representation as there can be many objects that are, say, homomorphic to one another without one being a representation of the other. It is only with respect to the first objection —that isomorphism has the wrong formal properties— that other mappings fare better because they can evade some of isomorphism's difficulties (e.g. embeddings need not be symmetric). But this improvement is not sufficient to compensate for all other difficulties and so I conclude that they do not provide us with a satisfactory answer to the enigma. And the same goes for the problem of style. As isomorphism, they can be a good answer to the factual variant of the problem but it does not seem to be the case that all scientific representations conform to one of these patterns.

According to an alternative version of the semantic view, the relation between a model and its target is similarity rather than isomorphism (Giere 1988, Ch. 3; 1999; 2004). Accordingly we obtain: model M represents target system T iff M is similar to T .

This view imposes fewer restrictions on what counts as a scientific representation than the structuralist conception. First, it enjoys the advantage over the isomorphism view that it allows for models that are only approximately like their targets. Second, the similarity view is not committed to a particular ontology of models. Unlike the isomorphism view, it enjoys complete freedom in choosing its models to be whatever it wants them to be.

However, these advantages notwithstanding, the similarity view does not offer satisfactory answers to the above questions.

The problems similarity faces when understood as a response to the enigma by and large parallel those of isomorphism. It also has the wrong logical properties and it is

not necessary for representation. As isomorphism claims, similarity claims rest on descriptions, but for a different reason. In saying that M resembles T one gives very little away. It is a commonplace observation that everything resembles everything else in any number of ways (see Goodman 1972). The claim that M is similar to T remains empty until relevant respects and degrees of similarity have been specified, which we do with what Giere (1988, 81) calls a ‘theoretical hypothesis’, a linguistic item.

Similarity *per se* does not provide us with a satisfactory answer to the problem of style either. An unqualified similarity claim is empty; relevant respects and degrees need to be specified to make a similarity claim meaningful. So what we need is an account of scientifically relevant kinds of similarity, the contexts in which they are used, and the cognitive claims they support. Before we have specifications of that sort at hand, we have not satisfactorily solved the problem of style in either its normative or its descriptive variant.

What about Giere’s ontological claim that models are abstract entities (1988, 81)? It is not entirely clear what Giere means by ‘abstract entities’, but his discussion of mechanical models suggests that he uses the term to designate fictional entities. To regard models as fictional entities is an interesting suggestion, but one that is in need of qualification. Fictional entities are beset with difficulties and in the wake of Quine’s criticisms most analytical philosophers have adopted deflationary views. Can fictional entities be rendered benign, and if so how exactly are they used in science? This is an interesting and important problem but, as Fine (1993) has pointed out, one that has not received the attention it deserves.

Let me conclude this section with some remarks on accounts of modelling other than the ones suggested within the framework of the semantic view of theories. During the last two decades a considerable body of literature on scientific modelling has grown and one might wonder whether this literature bears answers the question that I have been raising in this paper. In the case of the enigma and the ontological puzzle this does not seem to be case. The questions of where the representational power of models comes from and what kind of objects models are have not received much attention.²² With regards to the problem of style the situation is somewhat different. Debates over the nature of idealisation and the functioning of analogies have been prevalent for many years, and these can be understood as at least partially addressing the problem of style. The problem with the issue of style is a lack of systematisation rather than a lack of attention. Icons, idealisations and analogies are not normally discussed within one theoretical framework. As a consequence, we lack comparative categories that could tell us what features they share and in what respects they differ. What we are in need of is a systematic enquiry, which provides us with both a characterisation of individual styles and a comparison between them.

²² Noteworthy exceptions are Bailer-Jones (2003) and Suárez (2004); but no full-fledged account of representation has emerged yet from these discussions.

9. Outlook

In sections 5 and 6 I have argued that structural claims rest on more concrete descriptions of the target system. For this reason, descriptions are an integral part of any workable conception of scientific representation and we cannot omit them from our analysis. This is more than a friendly, but slightly pedantic and ultimately insignificant amendment to the structuralist programme; it casts doubt on a central dogma of the semantic view of theories, namely that models are non-linguistic entities. Models involve both non-linguistic and linguistic elements.²³

If I am right on this, the face of discussions about scientific representation will have to change. In the wake of the anti-linguistic turn that replaced the syntactic view with the semantic view of theories questions concerning the use of language in science have been discredited as misguided and obsolete. This was too hasty a move. There is no doubt that the positivist analysis of theories is beset with serious problems and that certain non-linguistic elements such as structures do play an important role in scientific representation; but from this it does not follow that language *per se* is irrelevant to an analysis of scientific theories or models. Scientific representation involves an intricate mixture of linguistic and non-linguistic elements and what we have to come to understand is what this mixture is like and how the different parts integrate. What kinds of descriptions are employed in scientific representation and what role exactly do they play? What kinds of terms are used in these descriptions? These are but some of the questions that we need to address within the context of a theory of scientific representation. And this also seems to tie-in nicely with the conclusion of section 4, because the intentionality required for scientific representation seems to enter the scene via the descriptions scientists use to connect structures to reality.

A sceptic might reply that although there is nothing wrong with my claim that we need descriptions, there is not much of an issue here because what we are ultimately interested in is the isomorphism claim itself and that such a claim is made against the background of some description may be interesting to know but is ultimately insignificant. I disagree. Neat phrases like '*S* is isomorphic to *T* with respect to description *D*' are deceptive in that they make us believe that we understand how the interplay between structures, descriptions and the world works. This is wrong. These expressions are too vague to take us anywhere near something like an analysis of scientific representation and more needs to be said about how structures, targets and descriptions integrate into a consistent theory of representation.

Acknowledgements

I would like to thank Nancy Cartwright, José Diez, Stephan Hartmann, Carl Hofer, Paul Humphreys, Moshé Machover, Julian Reiss, Sally Riordan, Juha Saatsi, Mauricio Suárez, and Micheal Redhead for helpful discussion and/or comments on earlier drafts. I am also grateful to the audiences at the Joint Session of the Mind Association

²³ Chakravartty (2001) has come to a very similar verdict, although based on a different argument.

and the Aristotelian Society, the Universidad Complutense de Madrid and the University of Leeds for lively and stimulating discussions. Thanks to Andrew Goldfinch for his help with the preparation of the final version of this paper.

REFERENCES

- Balzer, Wolfgang, C. Ulises Moulines, and Joseph D. Sneed (1987). *An Architectonic for Science. The Structuralist Program*. Dordrecht: Kluwer Academic Publishers.
- Bogen, James, and James Woodward (1988). 'Saving the Phenomena', *Philosophical Review* 97, 303-52.
- Boolos, George S., and Richard C. Jeffrey (1989). *Computability and Logic*. 3rd ed., Cambridge: Cambridge University Press.
- Bueno, Otávio (1997). 'Empirical Adequacy: A Partial Structure Approach', *Studies in the History and Philosophy of Science* 28, 585-610.
- (1999). 'What is Structural Empiricism? Scientific Change in an Empiricist Setting', *Erkenntnis* 50, 59-85.
- Cartwright, Nancy (1999). *The Dappled World. A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.
- Chakravartty, Anjan (2001). 'The Semantic or Model-Theoretic view of Theories and Scientific Realism', *Synthese* 127, 325-45.
- Da Costa, Newton C., and Steven French (1990). 'The Model-Theoretic Approach to the philosophy of Science', *Philosophy of Science* 57, 248-65.
- (2000). 'Models, Theories, and Structures: Thirty Years On', *Philosophy of Science* 67, Proceedings, 116-127.
- Dummett, Michael (1991). *Frege. Philosophy of Mathematics*. London: Gerald Duckworth.
- Fine, Arthur (1993). 'Fictionalism', *Midwest Studies in Philosophy* 18, 1-18.
- French, Steven (1999). 'Models and Mathematics in Physics: The Role of Group Theory', in Jeremy Butterfield and Constantin Pagonis, *From Physics to Philosophy*. Cambridge: Cambridge University Press, 187-207.
- (2000). 'The Reasonable Effectiveness of Mathematics: Partial Structures and the Application of Group Theory to Physics', *Synthese* 125, 103-20.
- (2003). 'A Model-Theoretic Account of Representation', *Philosophy of Science* 70, 1472-1483.
- , and James Ladyman (1999). 'Reinflating the Semantic Approach', *International Studies in the Philosophy of Science* 13, 103-121.
- Giere, Ronald (1988). *Explaining Science*. Chicago: University of Chicago Press.
- (1999). 'Visual Models', in Ronald Giere, *Science Without Laws*. Chicago: University of Chicago Press, 118-146.
- (2004). 'How Models Are Used to Represent Physical Reality', *Philosophy of Science* 71, 742-752.
- Goodman, Nelson: (1968). *Languages of Art*. 2nd ed. Indianapolis and New York: Hackett Publishing, 1976.
- (1972). 'Seven Strictures on Similarity', in Nelson Goodman, *Problems and Projects*. Indianapolis and New York: Hackett Publishing, 437-446.
- Hellman, Geoffrey (1989). *Mathematics without Numbers: Towards a Modal-Structural Interpretation*. Oxford: Clarendon Press.
- Hofbauer, Joseph, and Karl Sigmund (1998). *Evolutionary Games and Population Dynamics*. Cambridge: Cambridge University Press.
- Hughes, R. I. G. (1997). 'Models and Representation', in: *Philosophy of Science* 64, Proceedings, 325-336.
- Kroes, Peter (1989). 'Structural Analogies between Physical Systems', *British Journal for the Philosophy of Science* 40, 145-54.
- McAllister, James W. (1997). 'Phenomena and Patterns in Data Sets', *Erkenntnis* 47, 217-228.

- Morgan, Marry S., and Margaret Morrison (1999). 'Models as Mediating Instruments', in Marry S. Morgan and Margaret Morrison (eds.), *Models as Mediators. Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press, 10-37.
- Mundy, Brent (1986). 'On the General Theory of Meaningful Representation', *Synthese* 67, 391-437.
- Newman, M. H. A. (1928). 'Mr. Russell's "Causal Theory of Perception"', *Mind* 37, 137-148.
- Psillos, Stathis (1999). *Scientific Realism. How Science Tracks Truth*. London: Routledge.
- Redhead, Michael (2001). 'The Intelligibility of the Universe', in Anthony O'Hear (ed.), *Philosophy of the New Millennium*. Supplement to *Philosophy*, Royal Institute of Philosophy Supplement 48, 2001, 73-90.
- Resnick, Michael D. (1997). *Mathematics as a Science of Patterns*. Oxford: Oxford University Press.
- Rickart, Charles E. (1995). *Structuralism and Structure. A Mathematical Perspective*. Singapore: World Scientific.
- Russell, Bertrand (1919). *Introduction to Mathematical Philosophy*. London and New York: Routledge 1993.
- Schier, Flint (1986). *Deeper into Pictures. An Essay on Pictorial Representation*. Cambridge: Cambridge University Press.
- Shapiro, Stewart (2000). *Thinking About Mathematics*. Oxford: Oxford University Press.
- Stich and Warfield (1994). *Mental Representation. A Reader*. Oxford and Cambridge (MA): Blackwell.
- Suárez, Mauricio (2003). 'Scientific Representation: Against Similarity and Isomorphism', *International Studies in the Philosophy of Science* 17, 225-244.
- (2004), 'An Inferential Conception of Scientific Representation', *Philosophy of Science* 71, Supplement, 767-779.
- Suppe, Frederick (1989). *The Semantic Conception of Theories and Scientific Realism*. Chicago: University of Chicago Press.
- Suppes, Patrick (1960a). 'A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences' in Patrick Suppes, *Studies in the Methodology and Foundations of Science. Selected Papers from 1951 to 1969*. Dordrecht: Kluwer Academic Publishers 1969, 10-23.
- (1960b). 'Models of Data', in Patrick Suppes, *Studies in the Methodology and Foundations of Science. Selected Papers from 1951 to 1969*. Dordrecht: Kluwer Academic Publishers 1969, 24-35.
- (1970). *Set-Theoretical Structures in Science*. Stanford: Institute for Mathematical Studies in the Social Sciences, Stanford University.
- Swoyer, Chris (1991). 'Structural Representation and Surrogate Reasoning', *Synthese* 87, 449-508.
- Teller, Paul (2001). 'Whither Constructive Empiricism', *Philosophical Studies* 106, 123-150.
- van Fraassen, Bas C. (1980). *The Scientific Image*. Oxford: Oxford University Press.
- (1981). 'Theory Construction and Experiment: An Empiricist View', *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 2, 663-677.
- (1985). 'Empiricism and the Philosophy of Science', in Paul M. Churchland and Clifford Hooker (eds.), *Images of Science. Essays on Realism and Empiricism with a Reply from Bas C. van Fraassen*. Chicago and London: University of Chicago Press, 245-308.
- (1989). *Laws and Symmetry*. Oxford: Oxford University Press.
- (1991). *Quantum Mechanics*. Oxford: Oxford University Press.
- (1994). 'Interpretation of Science; Science as Interpretation', in Jan Hilgevoord (ed.), *Physics and our view of the World*. Cambridge: Cambridge University Press, 169-187.
- (1995). 'A Philosophical Approach to Foundations of Science', *Foundations of Science* 1, 5-9.
- (1997). 'Structure and Perspective: Philosophical Perplexity and Paradox', in M. L. Dalla Chiara et al. (eds.), *Logic and Scientific Methods*. Dordrecht: Kluwer Academic Publishers, 511-30.
- (2001). 'Structure: It's Shadow and Substance.' Forthcoming in P. Ehrlich and R. Jones (eds.), *Reverberations of the Shaky Game: Essays in Honor of Arthur Fine*. Chicago: University of Chicago Press.
- (2002). *The Empirical Stance*. New Haven and London: Yale University Press.
- (2004). 'Scientific Representation: Flouting the Criteria', *Philosophy of Science* 71, 794-804.
- Woodward, James (1989). 'Data and Phenomena', *Synthese* 79, 393-472.

Roman Frigg is a lecturer in philosophy at the London School of Economics. He received his PhD from the University of London. His main research interests lie in the field of general philosophy of science and philosophy of physics.

ADDRESS: Department of Philosophy, Logic and Scientific Method. London School of Economics. Houghton Street. London WC2A 2AE, England. E-mail: r.p.frigg@lse.ac.uk.

There Is No Special Problem About Scientific Representation¹

Craig CALLENDER and Jonathan COHEN

BIBLID [0495-4548 (2006) 21: 55; pp. 67-85]

ABSTRACT: We propose that scientific representation is a special case of a more general notion of representation, and that the relatively well worked-out and plausible theories of the latter are directly applicable to the scientific special case. Construing scientific representation in this way makes the so-called “problem of scientific representation” look much less interesting than it has seemed to many, and suggests that some of the (hotly contested) debates in the literature are concerned with non-issues.

Key words: scientific representation, mental representation, models.

... important philosophical problems concerning language have been misconstrued as relating to the content of science and the nature of the world.

van Fraassen 1980, 196.

1. Introduction

The harmonic oscillator, Ising model, and logistic map are typical representative structures used in science. In recent years the question of how such models can be about parts of the world has led to a burgeoning literature. Philosophers find it particularly puzzling how models, which commit sins of omission and commission by lacking and having features the world does and does not have, respectively, can nevertheless be about bits of the world. There are now a variety of different accounts of how scientific models represent, and of course, the usual philosophical squabbling over which one is right. It seems that a new philosophical problem has been discovered and philosophers of science have dutifully risen to the call.

Perhaps, however, they shouldn't have. For it is not clear that there is a special problem about scientific representation, as opposed to artistic, linguistic, and culinary representation. While philosophers have been quick to provide answers, few have spent time discussing the nature of the problem.

We'll undertake such an examination in this paper. We'll propose a more general framework in which to think about scientific representation that solves or dissolves the so-called “problem of scientific representation” while shedding light on many other questions surrounding scientific models. While the view we'll be advocating does not make all of the work on scientific representation insignificant, it does suggest that some of the debates in the literature are concerned with non-issues. Our framework re-orientates much of this work, so that some of it survives if understood as answering a different question than one about the nature of scientific representation *per se*.

¹ This work is fully collaborative; the authors are listed alphabetically.



2. *The Alleged Problem of Scientific Representation*

Current work on scientific representation is best appreciated against the backdrop of developments in philosophical conceptions of scientific theories beginning in the 1960s. In this work, Patrick Suppes and others developed the so-called semantic view of scientific theories, according to which the whole class of semantic or meta-mathematical models of the theory provides its semantic content (Suppes 1967, 1969; van Fraassen 1980; Giere 1988). Whatever its virtues and vices, the semantic view made newly salient the problem of explaining the relationship between models and the world.²

Whether one understands models as abstract or concrete, abstractions from theory or not, many philosophers have worried that they are not the sorts of things that are truth-apt, or even approximate-truth-apt. Just as there seems to be something wrong with claiming that a toy model airplane is true or false, there seems something wrong with claiming that an Ising model, Bloch model, or logistic map is true or false. Yet, even if models (unlike propositions, sentences, etc.) are not, or are not always, truth-apt, they are about the world in some sense. Surely it is correct to say that models can represent the world. This situation invites us to ask a question that has become one of the strands in the alleged problem of scientific representation, and that we shall call ‘the constitution question’: what constitutes the representational relation between a model and the world?

Various answers have been proposed to the constitution question. For example, Giere seems to be offering one in saying that there is a relation of “fit” or “similarity” to some degree and in some respects between a model and the world (Giere 1988, 81), where the respects and degree are picked out by scientists’ intentions in designing and using the model (Giere 1992, 122-123; but see note 7). Others instead think the relationship between model and the real world is one of isomorphism, partial isomorphism (French 2002), inference generation (Suárez 2003), and more.

In recent years, these issues have been woven closely together with related but distinct problems. For example, consider the “DDI” theory of representation of (Hughes 1997). Looking at Galileo’s use of a geometric figure in solving a problem in kinematics, Hughes argues that scientific representation typically has elements of “denotation” (elements of the model, e.g., lines, denote phenomena), “demonstration” (one uses the model to get a result) and “interpretation” (the result is then interpreted physically). DDI is not meant to provide necessary and sufficient conditions for when a representation takes place; rather Hughes is “making the more modest suggestion that, if we examine a theoretical model with these three activities in mind, we shall achieve some insight into the kind of representation that it provides” (Hughes 1997, S329; see also S335).

Here it seems that Hughes is interested in distinguishing scientific from other sorts of representation —i.e., he is attempting to solve a kind of demarcation problem for

² This problem can also be raised for those “mediating models” theorists who hold that scientific models are to some extent independent of theory.

scientific representation. He claims that DDI will inform us (typically) about what kind of representation occurs; for example, it will distinguish (i.e., demarcate) Galileo's scientific scribbles from Vermeer's masterly strokes. Hughes also criticizes Giere's "similarity" theory by pointing out the seeming truism from Goodman (1976) that every pair of entities is similar in some respects and dissimilar in others. Since the 'D' of denotation is set against Giere's similarity account, it is tempting to conclude that at least one 'D' is supposed to be a solution to Giere's problem (see especially pp. 6-8; see also Hughes 1999, 126). As we read him, then, Hughes offers his DDI proposal as an answer to both the constitution question and the demarcation problem about representation.

The issues are complicated further by another problem that comes out in Morrison's characterization of "the heart of the problem of representation" as the question "in virtue of *what* do models represent and how do we identify what constitutes a correct representation?" (Morrison 2006). There are at least two distinct questions here. The 'in virtue of what' question clearly sounds like the constitution question we took Giere to be addressing. But the second half of the quotation introduces a distinct problem: the normative issue of what it is for a representation to be correct. Many writers in the "models as mediators" school have focused on the normative question of what makes some models explanatory (cf. Morrison and Morgan 1999). Morrison (1999) claims that the representational and explanatory capacities of a model are interconnected (40). Inasmuch as 'interconnected' means that the explanatory/normative questions presuppose answers to the constitutive ones, we agree. But we do not believe the two questions are any more deeply connected.

Our feeling is that many authors writing on models don't contrast these questions as sharply as they should. For example, Bailer-Jones (2003) demands an answer to the constitution problem, and criticizes DDI for failing to provide such an answer. But it is not clear that DDI is intended as an answer. The inference generation theory of Suárez (2003) is explicitly directed at the constitution question, but it is not clear that the views Suárez criticizes (e.g., that of van Fraassen) are directed at the same question. Other work—for instance, much of that featured in Morrison and Morgan (1999)—seems focused on the normative problem. Still others, e.g., Hughes, also want to tackle the demarcation problem. In our view, running these issues together is conducive to confusion.

We would be remiss if we didn't mention that much of the writing about models concentrates on the fact that models misrepresent in some respects. How can they represent if they, well, mis-represent? For instance, the Lorentz model of convection in the atmosphere misses out on a variety of features of the Earth's real convection patterns; the model ignores scores of parameters relevant to the atmosphere and makes a number of false assumptions. It only captures a very small piece of the dynamical behavior of air (see Smith 1998, 9-13). As another example, consider that a Hardy-Weinberg model of a rabbit population will assume there are an infinite number of rabbits (to rule out the possibility of genetic drift). That's a lot of rabbits. As a

limiting case, many have worried that some models, like Bohr's of the atom, seem in some sense inconsistent (French 2002).

These kinds of problems have led philosophers to what we consider some pretty desperate measures. For example, Stephen French, a would-be defender of isomorphism, has retreated to the weaker claim that models must be partially isomorphic to the real world if they are to represent (French 2002). Likewise, the idealization and abstraction of models leads Bailer-Jones to the proposal that models entail certain propositions in some non-logical (and, as far as we can tell, magical) sense. An architect's plans for a bridge, just lying there on a desk, "entails" various propositions, according to her theory. Though her positive proposal is opaque to us, she is initially concerned with the ratio of true propositions to false propositions "entailed" by such objects as a way of saving the representational capacities of idealized models.

To those familiar with other theories of representation (and even those not), many of the concerns that seem to be driving such philosophical proposals may seem strange. To see why, consider a quotidian example of representation from outside science and notice how the questions analogous to those philosophers ask about scientific representation fail to get much of a grip. For example, consider the lowly stop sign. Are stop signs at intersections isomorphic or partially isomorphic to the imperative 'stop!' that they represent? Do they non-logically entail more true propositions than false ones? Taking another example, do the marks 'cat' in any way resemble real cats? Are philosophers of language worried that the marks 'cat' aren't furry or that cats lack constituents that are part of an alphabet? These questions about non-scientific representations strike us as bad ones, and we hope they strike you that way too. This suggests to us that there may well be something wrong with the questions being asked about scientific representation. Therefore, before further answers are given, we think it is high time to think a bit about what the questions are supposed to be.

3. *Scientific Representation, Meet Philosophy of Mind*

How are philosophers to understand scientific representation? Three *prima facie* plausible observations can guide us.

The first is that, in general, it is economical and natural to explain some types of representation in terms of other, more basic types of representation. We'll call this idea 'General Griceanism', as it amounts to a generalization of Grice's important views on representation. The General Gricean holds that, among the many sorts of representational entities (cars, cakes, equations, etc.), the representational status of most of them is derivative from the representational status of a privileged core of representations. The advertised benefit of this General Gricean approach to representation is that we won't need separate theories to account for artistic, linguistic, representation, and culinary representation; instead, the General Gricean proposes that all these types of representation can be explained (in a unified way) as deriving from some more fundamental sorts of representations, which are typically taken to be mental states. (Of course, this view requires an independently constituted theory of representation for the fundamental entities.)

The second observation is that, so long as we are in the General Gricean business of describing dependency relations among various sorts of representations, there is good reason to think that we should extend this treatment to scientific representations—i.e., that we should take the latter to be located somewhere in the web of dependency relations with other types of representations. After all, scientists routinely use entities other than models—language, pictures, mental states, and so on—to represent the very same targets that models represent. This coincidence of representational targets is explicable if (i) scientific representations get their representational status from linguistic (etc.) representations, or (ii) vice versa, or (iii) scientific representations and linguistic (etc.) representations get their representational status from some third sort of representation. But it would be surprising that scientific, linguistic, pictorial, mental, and other sorts of representations should coincide in their representational targets were they not at all related in the way that the General Gricean treatment indicates that they should be.

The third and final observation is that, if we distinguish derivative from fundamental representations, and are attempting to include scientific representations in the mix, it is reasonable to think that they belong among the derivative representations rather than the fundamental ones. For one thing, the distinction between science and non-science is famously elusive. Does Freud’s model of the unconscious represent one way if Freud’s theory is scientific and another (derivative) way or not at all if not? That seems unsatisfactory. Whether and how the model is about something shouldn’t hang on this classification.

Now, General Griceanism is so general, as stated, that discussion of it is much more easily carried out by reference to Grice’s specific version of the position that sometimes goes under the label ‘intention-based semantics’, and that we’ll call ‘Specific Griceanism’ in order to contrast with General Griceanism. We’ll advert to Specific Griceanism at times in what follows partly just to facilitate discussion, and partly as a way of showing that, since there are proposals about how to fill in the details, General Griceanism is not a mere promissory note. Despite this policy, however, we don’t want to be committed too much to the (by comparison, controversial) details of Specific Griceanism, and so present the latter only as an example.

3.1. Explaining Representation

As noted, the General Gricean proposes to distinguish between fundamental and non-fundamental representation, and to explain the latter in terms of the former. The Specific Gricean version of this distinction is made between so-called natural representation and non-natural representation. Natural representations are those whose representational powers are constituted independently of the mental states of their users/makers; these would include the number of rings on a tree (representing the age of the tree), the presence of smoke (representing the concomitant presence of fire), and so on. Non-natural representations, by contrast, are produced by human beings for the purpose of communicating something to an audience; this class would include linguistic tokens, some artworks, pre-arranged signals, and the like. To a zeroth ap-

proximation, the Specific Gricean program attempts to explain representation by giving a reductive account of non-natural representation in terms of natural representation. The next step (about which Grice himself had relatively little to say) is to combine the latter reduction with a naturalistic, reductive account of natural representation, thereby providing a full, naturalistically acceptable, reductive account of representation.

At the risk of obscuring the generality of General Griceanism, it may help to consider the Specific Gricean explanation of linguistic representation. Grice clearly thinks linguistic tokens are non-natural representations, so he proposes to use the general strategy outlined above to explain what he calls ‘speaker meaning’ —i.e., what it is for a speaker *S* to mean something by uttering *U* in terms of his acting with the intention of producing a belief or action in a hearer *H*. That is, he hoped to give a theory of roughly this form:

In uttering *U*, *S* means that *p* iff, for some *H*, *S* utters *U* intending in way ... to activate in *H* the belief that *p*.

Of course, the details of this Specific Gricean theory schema for speaker meaning are not without controversy (see Schiffer 1987, chapter 9). But the hope is that the theory will reduce the notion of speaker meaning for linguistic tokens to specific mental states of producers/hearers of these tokens —namely, the states of *S*’s intending to do something, and *H*’s believing that something else.

But the Specific Gricean’s job is not finished until she provides an account of the representational contents of mental states. This question about the metaphysics of representation for the fundamental units of representation is currently the subject of intense philosophical controversy. However, there is a range of popular answers to the question that are available for use at this stage of the Gricean explanation.³

There are several points about the Specific Gricean explanation of the representational powers of linguistic tokens that bear emphasis, and that provide lessons for General Griceanism, once we abstract away from the Specific Gricean details. We pause to belabor them.

First, notice that the account divides naturally into two stages. The first stage of Specific Griceanism consists in explaining the representational powers of linguistic tokens in terms of the representational powers of something more fundamental —namely, mental states. In the second stage, the Specific Gricean needs some other story to explain representation for the fundamental bearers of content, mental states.

³ Some of the most popular accounts of representation for mental states are functional role theories, informational theories, and teleological theories. A useful anthology is Stich and Warfield 1994; see also Cohen 2004 for a critical overview of much of this literature.

There is another (currently less popular) family of views of representation for mental states that should be mentioned —views according to which a mental state represents by virtue of being similar to its target in the sense that it occupies a similar position in an abstract phase space (cf., Churchland 1986; for criticism see Fodor and Lepore 1992, ch. 6). If something like this were correct, this would require some qualifications to some of our claims about the impotence of similarity in the constitution of scientific representation. We’ll return to this in note 9.

Likewise, the General Gricean view consists of two stages. First, it explains the representational powers of derivative representations in terms of those of fundamental representations; second, it offers some other story to explain representation for the fundamental bearers of content. Still, General Griceanism doesn't insist on the Specific Gricean way of drawing the line between its two stages.

Second, it is worth noting that, of these stages in either Specific or General Griceanism, most of the philosophical action lies at the second. The first stage amounts to a relatively trivial trade of one problem for another: you thought you had a problem of representation for linguistic tokens (or whatever you take to be derivative representations)? exchange it for a problem of representation for mental states (of whatever you take to be fundamental representations). This trade, in effect, just pushes back the problem of representation by a single step. The second stage, in contrast, amounts to a fairly deep metaphysical mystery. What is needed to solve it is a fundamental, non-derivative account of the metaphysics of representation; in particular, here it won't do to push the problem back a step. Accordingly, here there is sharp controversy surrounding matters large and small.

The third point is that the explanatory pattern at work here is extremely general. In particular, if you are sympathetic to this account of representation for linguistic tokens, you can use the same apparatus to generate accounts of representation for all sorts of other non-natural representations. For example, the very same apparatus answers this deep question about representation: how did the placement of a pair of lanterns in Boston's North Church belfry arch represent to Paul Revere that the British were coming by sea rather than land?⁴ Presumably Revere and the friend who sent him the signal, Joseph Warren, met beforehand and brought into being (by stipulation) their famous code: one if by land, two if by sea. Consequently, when Warren later determined that the British were indeed traveling by sea rather than land, he could reasonably intend that his hanging the pair of lanterns in the belfry would activate in his audience (Revere) the belief that the British would take the sea route. In this case, too, the initial question about representation (how does a pair of lanterns hanging in a belfry represent) is reduced, by a relatively trivial move, to a more fundamental question about how mental states represent. Having this one explanatory strategy, then, means having an account of representation that works for all sorts of representational objects (other than mental states, for which some other story about representation is needed).⁵

Fourth, as a reflex of its generality, the explanatory strategy we are now considering places almost no substantive constraints on the sorts of things that can be representational relata. Can the salt shaker on the dinner table represent Madagascar? Of

⁴ Our historical scholarship regarding this case was exhausted by consulting Longfellow's poem, "Paul Revere's Ride".

⁵ In yet another application, Fodor (1993) extends the same explanatory framework to the problem of the representational power of artworks, and uses this account to distinguish artworks from both rhetorical devices (say, the *Mona Lisa* from a shampoo advertisement) and mere things (say, Warhol's *Brillo Boxes* from Brillo boxes).

course it can, so long as you stipulate that the former represents the latter. Then, when your dinner partner asks you what is your favorite geographical land mass, you can make the salt shaker salient with the reasonable intention that your doing so will activate in your audience the belief that Madagascar is your favorite geographical land mass (obviously, this works better if your audience is aware of your initial stipulation; otherwise your intentions with respect to your audience are likely to go unfulfilled). Can your left hand represent the Platonic form of beauty? Of course, so long as you stipulate that the former represents the latter. Then, when your dinner partner asks you what you are thinking about, you can direct attention to your left hand with the reasonable intention that your doing so will activate in your audience the belief that you were thinking about the Platonic form of beauty. On the story we are telling, then, virtually anything can be stipulated to be a representational vehicle for the representation of virtually anything (including itself, in the odd circumstance where that is desired); the representational powers of mental states are so wide-ranging that they can bring about other representational relations between arbitrary relata by dint of mere stipulation. The upshot is that, once one has paid the admittedly hefty one-time fee of supplying a metaphysics of representation for mental states, further instances of representation become extremely cheap.

Fifth, the Gricean story we are telling allows for two distinct but related sorts of representation, examples of both of which have already come up in our discussion. On the one hand, there is representation of things (/properties/events/processes/etc.); thus, for example, a left hand can represent the family cat. On the other hand, there is representation of facts (/propositions/states of affairs/etc.); thus, for example, a left hand can represent that the family cat is on the mat. These two sorts of representation fit neatly into the same General Gricean explanation; in each case, the story is that the left hand represents what it does (a cat, a fact about a cat) by virtue of (i) an analogous representational relation that obtains between a mental state and its object (alternatively, a cat or a fact about a cat), together with (ii) a stipulation that confers upon the left hand the representational properties of that mental state. Indeed, the easy adaptability of the Gricean story to these different sorts of representation is a mere corollary of its indifference to the kinds of things that serve as representational relata. As noted, because our story puts almost no substantive constraints on the representational relata, it is neutral between representation of (or by) concreta and abstracta, the large and the small, and the near and the distant. The present point is just that the account is similarly neutral between representation of objects and facts.

Sixth, despite what was just said about the absence of constraints on the representational relata, there are plausibly pragmatic constraints on which representational vehicles and targets are used in particular cases. For example, the intentions underpinning the representational powers of salt shakers, left hands, and the like, are likely to go unfulfilled in the absence of certain kinds of communication. We take this consideration not to show that salt shakers and left hands are incapable of serving as full-blooded representational vehicles in principle. Rather, it shows that these objects, while capable of serving as full-blooded representational vehicles in principle, may not

do so in practice because they fail to serve the purposes at hand, given pragmatic constraints in force.

3.2. *Explaining Scientific Representation*

Our proposal, which will come as no surprise, is that scientific representation is just one more special case of derivative representation, and as such can be explained by the General Gricean account sketched above.

In particular, we propose that the varied representational vehicles used in scientific settings (models, equations, toothpick constructions, drawings, etc.) represent their targets (the behavior of ideal gases, quantum state evolutions, bridges) by virtue of the mental states of their makers/users. For example, the drawing represents the bridge because the maker of the drawing stipulates that it does, and intends to activate in his audience (consumers of the representational vehicle, including possibly himself) the belief that it does.

One might reasonably ask at this point why scientific representation could possibly be as useful and interesting as it undoubtedly is, were our analysis correct. Why bother to construct the drawing if its representational relation to the bridge is a product of mere stipulative fiat? Moreover, if fiat would as easily connect the bridge with anything at all, why not use cheaper (more readily available, more easily constructed) materials? In our view, the answers to these questions about scientific representations are no different from the answers to analogous questions about non-scientific representations. Just as the salt shaker (or, for that matter, the linguistic token ‘Madagascar’) is worth having for facilitating conversation about Madagascar in the absence of Madagascar, the drawing might be useful for facilitating conversation about the bridge in the absence of the bridge. Just as an upturned right hand is worth having because the geometrical structure it shares with the state of Michigan supports inferences about the geography of that state, the drawing of the bridge might (by virtue of preserving certain structural relationships among the represented parts) support inferences about the structure of the bridge.

But note that, just as in the case of similar questions about non-scientific representations, the questions about the utility of these representational vehicles are questions about the pragmatics of things that are representational vehicles, not questions about their representational status *per se*. Thus, if the drawing or the upturned right hand should happen not to rank highly along the dimensions of value considered so far, this would, on our view, make them non-useful vehicles that do represent, rather than debar them from serving as representational vehicles altogether.⁶

⁶ The idea that virtually anything can serve as a vehicle for scientific representation has met with some resistance, even scorn, in the literature (despite having been occasionally endorsed by some, e.g., Teller (2001, 397)). French writes “Not anything can serve as a scientific model of a physical system; if the appropriate relationships are not in place between the relevant properties then the ‘model’ will not be deemed scientific” (French 2002, 6). Bailer-Jones, in criticizing Hughes, points out that on Hughes’s account representation is stipulative, “as if ‘what represents what’ *could* be entirely arbitrary and merely set per decree. This could in some instances preclude that a model is about the empirical world in any meaningful and informative way” (Bailer-Jones 2003, 72).

Presumably scientific contexts come with their own set of pragmatic constraints, and these may drive the choice among possible scientific representations in ways that are idiosyncratic to science. For example, pathological cases like Weierstrass's example of a continuous but nowhere differentiable function ($f(x) = \sum_{k=1}^{\infty} [(\sin(\pi k^a x)) / (\pi k^a)]$) will not typically be used in science, nor would scientists use the picture of people climbing up a growth chart from the Microsoft clip-art that comes with every PC, or live jellyfish. And we can make conjectures at (and, in principle, even investigate) the reasons for these constraints. Weierstrass's pathological function typically won't be the first choice for scientific representation because scientists usually want to use the functions they choose, and that usually means differentiating them. The silly pre-drawn Microsoft graph that comes with most PCs, by contrast, won't be used for sociological reasons: it would simply be too embarrassing to have a graph from a Microsoft picture gallery in an academic economics journal (on the other hand, it might be used to represent in a Powerpoint display in a business' human resources department). Finally, live jellyfish won't be used because they can sting.

That said, it should be clear that the constraints ruling out these choices of would-be representational vehicles are pragmatic in character: they are driven by the needs of the representation users, rather than by essential features of the artifacts themselves.

Likewise, we suggest that, while resemblance, isomorphism, partial isomorphism, and the like are unnecessary for scientific representation, they have important pragmatic roles to play; namely, they can (but need not) serve as pragmatic aids to communication about one's choice of representational vehicle.

To see this, consider again the problem first raised for the salt shaker—that of making one's representational stipulations clear to one's audience. One alternative to announcing the stipulated representational relationship is to make one's intentions obvious by choosing a representational vehicle (from among indefinitely many candidates) that resembles its representational target in salient respects. For example, the geometric similarity between the upturned human right hands and the geography of Michigan make the former a particularly useful way of representing relative locations in Michigan, and it normally would be foolish (but not impossible!) to use an upturned left hand for this purpose since a more easily interpreted representational vehicle is typically available. Similarly, the behavior of billiard balls may prove a useful choice of model for the behavior of elastic particle interactions in a gas because there is a salient similarity/isomorphism between the dynamics of the vehicle's objects (billiard balls) and the target's objects (gas particles). This is not to say that the very same target could not be represented by an upturned left hand, or anything else for that matter,

To our eyes, these sentiments seem motivated more by intuition than argument; we suspect they come from running together the constitution question (what constitutes representation?) with the normative question (what makes a representation a good one?). We propose that intuitions to the effect that such and such cannot serve as a model are best understood as reflecting the unlikelihood of anyone's using such and such as a model, given certain assumptions about pragmatic purposes. If so, then our view accommodates them.

but only that similarity/isomorphism can make one of these choices more convenient than the other (given the scientific purposes at hand).

Our proposal, then, is that scientific representation is just another species of derivative representation to which the General Gricean account is straightforwardly applicable. This means that, while there may be outstanding issues about *representation*, there is no special problem about *scientific* representation.

4. *Surrounding Problems Dissolved/Reframed*

Once our view of scientific representation is in place, the surrounding landscape of problems —problems that have inspired much of the philosophical interest in models— changes dramatically. This can be viewed in two ways. The more dramatic assessment would be to say that these problems have been dissolved. The less dramatic (but probably more accurate) assessment would be to say that our view allows for the fruitful reframing of these problems as pragmatic issues about which among alternative (and equally viable) representations best meet scientific needs.

4.1. *What Does it Take For x to Represent y ?*

We've seen that a cottage industry has arisen in recent years around what we called (in §1) the constitution problem about scientific representation: what does it take for x to constitute a scientific representation of y ? Some (French) hold that x and y must stand in some sort of isomorphism (or partial isomorphism), while others (Giere, Teller) insist that what is crucial to representation is that x is similar to y .⁷ Still others (Suárez) have argued that it is essential to representation that x allows its users to generate inferences about y . Suffice to say that the debates between proponents of these different accounts have not resulted in consensus. As far as we can see, all of the proposals are either vacuous or too demanding. Since there is always, trivially, some or other isomorphism of structure, similarity, or generated inference that relates an arbitrary x to an arbitrary y , the accounts in question will be vacuous if they are not supplemented with a robust account of what sort of isomorphism, what respect of similarity, or what sorts of inference generation, are required. On the other hand, it has proven exceedingly difficult to specify the needed sense of isomorphism, similarity, or inference gen-

⁷ However, it is possible to read Teller (2001) and Giere (1999) as appealing to similarity in a more deflationary way, and indeed in a way that ends up anticipating the position we are defending. For, while they claim that x represents y in virtue of a similarity between x and y , they also insist that there is no substantive sense of similarity that unites all vehicle, target pairs and that can be specified in advance. Rather, on their view, the relevant similarity relation is stipulated by users of the representations, according to their own purposes, on a case by case basis.

If this is right, then our disagreement with Teller and Giere is largely terminological. Our reason for preferring our own terminology is only that, insofar as the sense of 'similarity' is entirely given by stipulation on a case by case basis, it seems that representation is only nominally constituted by similarity. What does the real representational work, it turns out, is stipulation. Better, then, we think, to drop the empty talk of similarity in favor of an up-front admission that representation is constituted in terms of stipulation (plus an underlying account of representation for the mental states subserving stipulation), as per the General Gricean view we are defending.

eration in any detail: invariably, such specifications have been insufficiently general to cover the wide variety of instances of scientific representation.

From the perspective of the General Gricean story we've been telling, these difficulties are unsurprising. For if, as we've been urging, scientific (and other non-natural) representation is constituted in terms of a stipulation, together with an underlying theory of representation for mental states, isomorphism, similarity, and inference generation are all idle wheels in the representational machinery —none of them (on any understanding) amounts to a necessary condition on scientific, or any other non-natural, representation.

Is there, then, nothing at all to the traditional disputes over the role of isomorphism (etc.) in scientific representation? That seems to us not quite right either. We are not denying that isomorphism, similarity, and inference generation may relate representational vehicle and representational target in many cases of scientific (and other non-natural) representation. We claim that these conditions do not constitute the representational relation, and hence are not necessary features of representation. However, we allow that there are important roles for these conditions —viz., they may serve as pragmatic aids to the recognition of a representational relation that is constituted by other means.⁸ Moreover, since the expectations representation users have about how audiences will interpret form an important part of the story we're telling, such pragmatic aids can constrain our choices about which representational relations to use.

If this is right, then there will remain a role for considerations about isomorphism, similarity, and inference generation after all. Namely, these considerations (and possibly others) may contribute to an anthropology of the use of scientific representations by providing a taxonomy of the sorts of pragmatically guided heuristics scientists bring to bear on their choices between representational vehicles.⁹ But if so, then there is no longer any reason to think that there is a conflict between, say, Giere's similarity and Suárez's inference generation, and so no reason that there should be a dispute between proponents of such accounts: these are simply independent pragmatic constraints that may work together or separately to guide choices between scientific representations. This point, we think, should serve to undercut that growing proportion of the literature on scientific representation devoted to arguing in favor of one of these

⁸ In saying that the constraints on representational vehicles are pragmatic in character, we certainly don't mean to deny that they have epistemic force or rationale. On the contrary, it is plausible that the pragmatic constraints on scientific representation typically will center around epistemic demands insofar as scientists *qua* scientists are in the business of acquiring knowledge about the world.

⁹ Recall that, on some views, the fundamental level of representation appealed to by the General Gricean is itself constituted in terms of similarity. If some such similarity-based view were correct, this would mean that similarity has a role to play in the explanation of scientific representation that goes beyond the role we've allowed for it in the main text. On the other hand, even on the envisioned scenario, the relata related by similarity would be (not scientific models and worldly targets, but) mental states and worldly targets. Consequently, even this outcome would fail to give the defenders of similarity *qua* explanation of scientific representation what they most seem to want.

accounts and against the others; if, as we contend, these accounts are not in competition, this should spare the needless consumption of much ink and many trees.

4.2. How Do Models Represent Despite Idealization?

If our General Gricean story is correct, the question of how models can represent despite their use of idealization, abstraction, etc., can't really be a question of how they manage to represent.

It is important to be clear that one can succeed along the dimension of representation but fail along the dimension of truth: something can be a representation although it represents falsely or comes up short on various pragmatic measures. For example, suppose the instruction is "one if by land, two if by sea," and the British come by sea but Warren hangs only one lantern. Then Warren would have successfully induced in the mind of Revere the belief that the British are coming by land. The representation would have induced in Revere a false belief. It misrepresents (i.e., represents falsely) the situation to Revere; moreover, given that the point of Warren and Revere's coordination was to produce a true belief in Revere's head, the representation meets its goals badly. But, for all that, it is still a representation.

Clearly, as Morrison (2006) emphasizes, looking at the details of the model in isolation will not answer the question of whether it represents truly, falsely, or approximately truly. Truth, falsity, and approximate truth are features that putatively apply to things that are representations; as such, the question of whether x represents y is independent of (indeed, prior to) the question of whether x is a true, false, or approximately true representation of y . Contrary to what many seem to have thought, then, there is no reason for fearing that the merely approximate status of a model impugns its capacity to represent.

5. Objections and Replies

5.1. Whither Realism?

Our view is extremely permissive about representation —it requires only an act of stipulation to connect representational vehicle with representational target (once the underlying metaphysics of representation for mental states is in place). It is so permissive, in fact, that it might suggest that we have begged the question in favor of irrealism about the posits of science. After all, if all that is required is mere stipulation, there is nothing to distinguish a stipulation connecting a vehicle to electrons, on the one hand, from a stipulation connecting a vehicle to phlogiston, on the other. But, a realist would say, this is a distinction we really want to make between representations, insofar as the former model tells us something about what really exists in the world (electrons) and the latter model tells us something about what really does not (phlogiston).

We would indeed be bothered if our view of scientific representation precluded realism about the posits of science. For one thing, we are rather fond of electrons. Moreover, we would strongly prefer not to have our commitments about the realism-antirealism debate decided by our theory of representation.

Luckily, there is no clash between realism and our view of scientific representation. For, while our approach to the question of what constitutes representation marks no distinction between models of existent entities and models of non-existent entities, it leaves plenty of room for further distinctions between such models, including those that the realist needs to get her position off the ground. In particular, the relevant distinction between a model of electrons and a model of phlogiston is not that one counts as a model and the other doesn't, but that one may be a better model in some normative respects than the other. Realists have famously offered a number of criteria—predictive and explanatory success, coverage of a wide range of data, etc.—that they use to measure the merit of models. Since, as we have emphasized, the constitution question about models is not identical to the various normative questions about the merits of things that are models, our proposal about the former question leaves room for many answers—including realist answers—to the latter questions.¹⁰

5.2. *Whither Irrealism?*

In responding to worries from realists (§5.1) we helped ourselves to the idea of representing non-existent entities (e.g., phlogiston). But how, one might ask, is representation of the non-existent possible on our view? After all, we have insisted that scientific representation is a relation between a representational vehicle and a representational target. And, insofar as relations cannot hold in the absence of relata, this commitment might seem to rule out the possibility that there is no genuine worldly entity that scientific models succeed in representing. For consider: if kicking is a relation, then you can't kick x unless x exists; if kissing is a relation, then you can't kiss x unless x exists (cf. Quine 1956). Likewise, if representation is a relation, then a model can't represent DNA unless DNA exists. But, just as we are loath to rule out realism about the posits of science, we would be equally embarrassed if our view about scientific representation ruled out irrealism about the posits of science.

This worry is too general to be a particular problem for us. First, the worry arises for anyone who thinks of scientific representation as any kind of relation at all. This crowd is broad indeed, and certainly includes all of the defenders of answers to the constitution problem that are competitors to our view: similarity, isomorphism, and related notions are proposed as ways of understanding representation as a relation too, so defenders of these views also owe a story about how we manage to represent the non-existent. Second, the worry arises for all species of representation—not just scientific representation—and there is no reason to suspect that whatever ultimately ex-

¹⁰ There are multiple normative dimensions along which models can be measured. Our claim is that models of electrons are good and models of phlogiston are bad along dimensions that realists have stressed as ways of distinguishing between the posits we should accept and the posits we shouldn't accept. But one might also be interested in independent dimensions of evaluation: e.g., how well the model communicates what one wants to communicate about the representational target, how well the model functions in science (e.g., does it help explain the phenomena?), etc.

plains representation of unicorns and golden mountains won't work for representation of phlogiston and the ether.¹¹

5.3. *Is the Cure More Controversial than the Sickness?*

In the foregoing we have made blithe use of the Gricean framework for explaining representation. On the other hand, there are outstanding difficulties in the details of that framework (for examples of disputes about these difficulties, see Grice 1989, ch6, ch14; Schiffer 1972; and Loar 1981). But if the details of the Gricean framework are not understood either, why think that appealing to them will shed any light on problems in the philosophy of science? Why hope to explain one mystery by appeal to another?

We have two main reasons for not being bothered by such outstanding controversies. First, everyone needs an account of mental representation. We should think philosophers would be delighted to learn that the price we are all already committed to paying in the philosophy of mind also buys a solution to the constitution problem about representation in the philosophy of science (even if no one has yet raised the funds). Second, as we pointed out when we introduced the Gricean apparatus in §3, our General Gricean proposal for understanding scientific representation is largely independent of the details of Specific Griceanism. Though we have appealed to Specific Griceanism in attempting to show that flesh can be put on the skeletal framework of General Griceanism, we do not mean to commit to Specific Gricean particulars. But, with one significant exception, the controversies about Grice's program are largely confined to the level of Specific Griceanism. Indeed, there seems to be a fairly solid consensus in favor of General Griceanism among the relevant experts in philosophy of mind and language, arguably for good reasons.

The significant exception to the idea that General Griceanism is insulated from controversy concerns the understanding of representation at the fundamental, non-derivative level. General Griceanism is, of course, committed to telling some story about the metaphysics of fundamental (typically mental) representation. However, while, as noted in §3.1, there is a notable absence of consensus even about the broad shape that a fundamental metaphysics of mental representation should take, nothing we have said chooses sides in these debates; consequently, General Griceanism is in trouble only if it turns out that mental representation is unreal.

In fact, even this commitment is dispensable; for, although in presenting General Griceanism we have welcomed the idea that the fundamental sort of representation at the bottom of the intentional stack is one that applies to mental states, it is easy to

¹¹ Thus having shifted the dialectical burden, we hasten to add that there are actual strategies for responding to the general worry, and that they seem applicable to scientific cases. First, one might bite the bullet and hold that, in cases where x doesn't exist, agents don't succeed in representing x but merely believe that they are representing x . Alternatively, one might appeal to a Humean strategy that (i) draws a distinction between atomic and compound representations, (ii) explains representation for the atoms by a relational theory, and (iii) explains compound representations as recursive structures built from other representations (cf. Hume 1777/1975, §II).

imagine a variant of General Griceanism that does without this idea. For purposes of illustration, suppose it turns out that, as urged by Dennett (1987), maybe Quine (1960; e.g., 221), Stich (1983), and others, mental representation is unreal. Then, so long as there is some genuine representation in the world by something not a scientific model—say, linguistic representation—we can still get our story off the ground by running a story analogous to Grice’s that construes scientific representation as derivative from this other sort of representation. So, really, the only way we can lose is if either (i) there is no representation anywhere, or (ii) scientific representation is fundamental. Both of these possibilities seem pretty far beyond the pale to us, so we aren’t particularly worried about them.

6. Conclusion: Where We Are Now

It is somewhat surprising that current disputes over scientific representation have often been carried out in isolation from more general work on representation. After all, this is justifiable only if scientific representation is constituted in a fundamentally different way from non-scientific representation, and that would seem to make a mystery of the possibility of expressing the content of scientific models by other means. Moreover, as we have shown, there are relatively well-worked out views about representation that seem to apply straightforwardly to scientific representation and substantially clarify the parochial disputes that have grown up around representation in philosophy of science.

Though we have deflated the constitution problem for models, there are still related questions that survive. We conclude by describing three of these questions and commenting on their relative interest.

First, we have claimed that anything can represent anything in science when the appropriate conditions are met. But what are these conditions and when are they typically met? Why, for example, did the Minkowski diagram triumph over the Loedel, Breheme and complex rotation diagrams as the standard vehicle of representing the spacetime of special relativity? How and why did the Feynman diagram come to dominate post-war physics (Kaiser 2000)? These anthropological questions, identified in §3, remain interesting questions in sociology of science.

Another question concerns the confirmatory and explanatory relationships between models, theories, and data. Though we would caution against overstating the “independence” of scientific models from overarching theories, one of the valuable lessons of the modeling literature is its study of the idea that models can sometimes take on a “life of their own” in science: the model can itself become the subject of scientific endeavor. We can think of no more prominent example than the Ising-Lenz model in statistical mechanics. When Onsager in 1944 ingeniously showed that the $d = 2$ Ising model displayed singular behavior despite having a non-vanishing partition function, he precipitated a real revolution in the study of phase transitions (see Domb 1996; Hughes 1999). For more than 50 years, large groups of physicists and mathematicians have devoted their time solely to solving various Ising models that represent an increasingly large number of systems.

Serious philosophical questions attend such changes. Suppose one asks how a physical system can exhibit multiple phases (solid-liquid-gas). Statistical mechanics answers by showing that the so-called Gibbs measure for the system is non-unique. But it demonstrates this non-uniqueness only for an infinite volume lattice with nearest-neighbor interactions, a simple interaction energy, and a host of other unrealistic assumptions. Physicists will want to say such a model represents the co-existence of phases in real (and hence finite volume non-lattice) macroscopic systems with complicated interactions. That such an idealized infinite system might represent a real macroscopic system is no problem. But physicists clearly also think that such a demonstration has important explanatory and confirmatory powers. Philosophers, however, might ask: Are the coexistence of phases in real systems explained by this model? Does this model confirm the basic tenets of statistical mechanics? Or of thermodynamics? And what relationship does this model have to the experimentally measured values of thermodynamic parameters of various gases? The question is similar to that asked recently about the explanatory/confirmatory status of computer-generated simulations. These simulations, like solutions to the Ising model, are often treated as having the status of genuine experiments. What are we to make of this in either case? The traditional questions of philosophy of science regarding explanation and confirmation arise again in the context of models. Much of the modeling literature has admirably examined the evidentiary/explanatory relationships between models that have a life of their own and theory and data, especially in various case studies. They have worked on what we called the normative problem in §2. By showing that there is no special puzzle about scientific representation, we hope to free these studies to focus on the confirmatory and explanatory role of models unencumbered by the perceived need to talk about the representation relation.

Finally, §2 described Hughes as seemingly interested in a kind of demarcation problem—that of saying what separates Galileo’s geometric figures from Vermeer’s masterpieces. Plausibly, scientific representation is just representation that takes place when the agents are scientists and their audiences are either fellow scientists or the world at large. But that means that to solve the demarcation problem in scientific representation one must first solve the prior question, THE demarcation problem famously discussed by Popper, Lakatos, Grünbaum, and Laudan. We are not optimistic about solving this problem. And we think it a virtue of our account that it allows one to see clearly that the demarcation problem for representation just is an instance of the general demarcation problem concerning the difference between science and non-science.

Demarcation worries aside, we’ve seen that there remain a number of interesting questions about representation in philosophy of science. We submit that ‘what constitutes scientific representation?’ is not one of them.¹²

¹² We are grateful to Nancy Cartwright, Paul Churchland, Andrew Hamilton, Sam Rickless, Mauricio Suárez, and Paul Teller, UCSD’s Philosophy of Science Reading Group, and an audience at the London School of Economics for helpful discussions of these matters.

REFERENCES

- Bailer-Jones, D. M. (2003). "When Scientific Models Represent," *International Studies in the Philosophy of Science* 17, 59-74.
- Churchland, P. M. (1986). "Some Reductive Strategies in Cognitive Neurobiology," *Mind* 95: 379, 279-309. Reprinted in Churchland (1989), 77-110.
- (1989). *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, MIT Press, Cambridge, Massachusetts.
- Cohen, J. (2004). "Information and Content," in L. Floridi (ed.), *Blackwell Guide to the Philosophy of Information and Computing*. New York: Basil Blackwell, 215-227.
- Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, Mass.: MIT Press.
- Domb, C. (1996). *The Critical Point: A Historical Introduction to the Modern Theory of Critical Phenomena*. Bristol, PA: Taylor & Francis, Ltd.
- Fodor, J. A. (1993). "Deja vu All Over Again: How Danto's Aesthetics Recapitulates the Philosophy of Mind," in M. Rollins (ed.), *Arthur Danto and His Critics*. Oxford: Blackwell, 41-54.
- , and E. Lepore (1992). *Holism: A Shopper's Guide*, Cambridge, Mass.: Blackwell.
- French, S. (2003). "A Model-Theoretic Account of Representation (Or, I Don't Know Much About Art... But I Know It Involves Isomorphism)," *Philosophy of Science Supplement* 70, 1472-1483.
- Giere, R. (1988). *Explaining Science: A Cognitive Approach*. Chicago: Chicago University Press.
- (1992). "Introduction: Cognitive Models of Science," in R. Giere (ed.), *Cognitive Models of Science*. Minneapolis: University of Minnesota Press, Vol. 15, xv-xxvii.
- (1999). "Using Models to Represent Reality," in L. Magnani, N. Nersessian, and P. Thagard (eds.), *Model-Based Reasoning and Scientific Discovery*. New York: Kluwer/Plenum, 41-57.
- Goodman, N. (1976). *Languages of Art: An Approach to a Theory of Symbols*. Indianapolis: Hackett.
- Grice, H. P. (ed.) (1989). *Studies in the Way of Words*. Cambridge, Mass.: Harvard University Press.
- Hughes, R. (1997). "Models and Representation," *Philosophy of Science* 64, S325-S336.
- (1999). "The Ising Model, Computer Simulation, and Universal Physics," in M. Morgan and M. Morrison (eds.), *Models as Mediators*. Cambridge: Cambridge University Press, 97-145.
- Hume, D. (1777/1975). *An Enquiry Concerning Human Understanding*. L. A. Selby-Bigge and P. H. Nidditch (eds.). Oxford: Oxford University Press.
- Kaiser, D. (2000). "Stick-Figure Realism: Conventions, Reification, and the Persistence of Feynman Diagrams, 1948-1964," *Representations* 70, 49-86.
- Loar, B. (1981). *Mind and Meaning*. Cambridge: Cambridge University Press.
- Morrison, M. (1999). "Models as Autonomous Agents," in M. Morgan and M. Morrison (eds.), *Models as Mediators*. Cambridge: Cambridge University Press, 38-65.
- (2006). "Models as Representative Structures," in S. Hartmann and L. Bovens (eds.), *Nancy Cartwright's Philosophy of Science*. Routledge. Forthcoming; available at http://www.uni-konstanz.de/ppm/workshop2/Morrison_final.pdf.
- , and M. Morgan (1999). "Models as Mediating Instruments," in M. Morgan and M. Morrison (eds.), *Models as Mediators*. Cambridge: Cambridge University Press, 10-37.
- Quine, W. V. O. (1956). "Quantifiers and Propositional Attitudes," *The Journal of Philosophy* 53, 177-187.
- (1960). *Word and Object*. Cambridge, Mass.: MIT Press.
- Schiffer, S. (1972). *Meaning*. Oxford: Clarendon Press.
- Schiffer, S. (1987). *Remnants of Meaning*. Cambridge, Mass.: MIT Press.
- Smith, P. (1998). *Explaining Chaos*. Cambridge: Cambridge University Press.
- Stich, S. (1983). *Form Folks Psychology to Cognitive Science: The Case Against Belief*. Cambridge, Mass.: MIT Press.
- , and T. A. Warfield (1994). *Mental Representation: A Reader*. Oxford: Basil Blackwell.
- Suárez, M. (2004). "An Inferential Conception of Scientific Representation," *Proceedings of the Philosophy of Science Association* 71: 5, 767-779.
- Suppes, P. (1960). "A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences," *Synthese* 24, 287-300.
- (1967). "What is a scientific theory?" in S. Morgenbesser (ed.), *Philosophy of Science Today*. New York: Basic Books, 55-67.

Teller, P. (2001). "Twilight of the Perfect Model Model," *Erkenntnis* 55, 393-415.
van Fraassen, B. (1980). *The Scientific Image*. Oxford: Clarendon Press.

Craig CALLENDER is a professor of philosophy at the University of California, San Diego. He has published widely on philosophy of science, philosophy of physics, and metaphysics, and especially on questions about spacetime, quantum mechanics, and the foundations of statistical mechanics.

ADDRESS: Department of Philosophy. University of California, San Diego. 9500 Gilman Drive. La Jolla, CA 92093-0119. E-mail: ccallender@ucsd.edu.

Jonathan COHEN is an associate professor at the University of California, San Diego. He works largely on topics in philosophy of mind, language, and perception, particularly as these are informed by the cognitive sciences. Much of his work in recent years has concerned color and color vision.

ADDRESS: Department of Philosophy. University of California, San Diego. 9500 Gilman Drive. La Jolla, CA 92093-0119. E-mail: joncohen@aardvark.ucsd.edu.