

Objective Similarity and Mental Representation

Alistair M. C. Isaac

June 6, 2012

Abstract

The claim that similarity plays a role in representation has been philosophically discredited. Psychologists, however, routinely analyze the success of mental representations for guiding behavior in terms of a similarity between representation and the world. I provide a foundation for this practice by developing a philosophically responsible account of the relationship between similarity and representation in natural systems. I analyze similarity in terms of the existence of a suitable homomorphism between two structures. The key insight is that by restricting attention to only those homomorphisms induced by causal processes, we can solve two philosophical problems with a single assumption. First, causal structure provides an adequate source for the bias required to ensure the similarity relation is non-trivial; second, it provides an adequate source for the directionality required to move from similarity to representation. I defend this account against objections by Goodman and van Fraassen and demonstrate that it is indeed the account of similarity's role in representation assumed by psychological practice.

1 Introduction

How do humans reason successfully about the world? A standard answer is that we *represent* the world and reason about it by manipulating these mental representations. But what special feature of a mental representation allows us to use it to reason about the world? It can't be the fact that it is a representation *simpliciter*. I can stipulate that this pencil represents Canada, but no amount of manipulation of the pencil will tell me the size of the population of Canada. So, what special feature makes a mental representation different from the pencil?

The classical answer to this question is that the mental representation is *similar* to the object it represents. Aristotle used the analogy of the signet ring to explain how this works. When a signet ring is pressed into warm wax, the wax receives the form of the ring, but not the matter. Aristotle argued that perceptual organs behave just like the wax. When the eye sees an object, it takes on its form, but not its matter. Since the mental representation has a similar form to the external object, it can be used to reason about that object (*De Anima*, Bk. II, Ch. 12).

From the perspective of philosophy, the similarity analysis of mental representation has fallen on hard times. Already in the seventeenth century, the idea that the properties of the mental and those of the physical are comparable was completely discredited. By the twentieth century, the idea that similarity could explain representation in any context was also undermined. Nelson Goodman argued forcefully against the value of similarity as a philosophical concept, exposing it as “a pretender, an imposter, a quack” (Goodman, 1970, p. 437).

Unfortunately, the psychologists don't seem to be listening. Implicit appeals to the structural similarity between a mental representation and the external object it represents appear repeatedly in research on perception and mental imagery (for summaries of some experimental findings, see e.g. Shepard, 1975, or Kosslyn et al., 2006). Even those few holdouts who reject the theoretical conclusions of the imagery literature (that there are distinctively picture-like representations) find themselves forced to allow for some kind of structural similarity between symbolic representation and object in order to account for the data (e.g. Pylyshyn, 2002).

The purpose of this paper is to reconcile psychological practice with philosophical theory. In particular, the task is to define a notion of objective similarity which can explain the informativeness of mental representations while avoiding Goodman-style counter-arguments. The key step is to allow ourselves the assumption of causal structure in the world. It turns out that if we assume the causes and categories of the current scientific view of the world, a natural notion of similarity falls out of them (Section 2).

The lynchpin of the argument that similarity cannot underwrite representation is directionality. Similarity is symmetrical, while representation is directed. Here again the assumption of causal structure can help. By combining the directionality of causal relations with objective similarity, we can produce a definition of representation in nature, or *N-representation*. N-representation is just another relation existing in nature, and does not depend in any way on human purposes (if it did, it would be explanatorily impotent for understanding mental representations). Van Fraassen has taken Goodman's arguments to imply that no such natural representational relation is possible, but we'll see that his argument only applies to a very restricted class of representations, not the concept in general (Section 3).

Finally, we'll conclude by surveying psychological theories of mental representation. We'll see that several prominent views take mental representations to be instances of N-representation. These internal N-representations graduate to full blown mental representations once they function to guide action. We'll conclude with a brief analysis of the source of *misrepresentation* on this account. The objective similarity between mental representations and objects in the world helps us to reason about the world, but it still leaves room for us to make mistakes (Section 4).

2 Objective Similarity

There are two natural strategies for defining similarity: property-based and isomorphism-based.¹ The property-based strategy takes two objects to be similar when they share properties. The isomorphism-based strategy takes two objects to be similar if a mapping between their component parts exists. The account developed here is an isomorphism-based account as this conforms more closely to the use of terms in the psychological literature. Nevertheless, it is worth stressing that the choice of properties or isomorphisms is essentially one of convenience. A property-based account can be recast in isomorphism terms, and an isomorphism account can be characterized in terms of shared second-order properties.

The fundamental challenge for any definition of objective similarity is accounting for bias without appealing to a human agent. After a survey of this problem, we'll examine some examples of similarity in nature, before introducing the full definition of objective similarity.

¹A note on terminology. I will use the term "isomorphism" for the general notion of a structure preserving mapping and "homomorphism" for the specific type of structure preserving mapping introduced in Section 2.3. Although somewhat mathematically lax, this choice follows the use of psychologists, who frequently use the term "isomorphism" when context makes clear that they mean "homomorphism" (as in the quotes from Gallistel and Shepard in Section 4.1).

2.1 The Challenge: *Whence the Bias?*

When applied naïvely, both property and isomorphism-based analyses of similarity suffer from exactly the same problem: *too many similarities*. In order to avoid this problem some kind of bias needs to be introduced. Without any source of bias, everything is similar to everything else, and similarity is a trivial notion. By privileging some properties (or mappings) over others, bias can reduce the number of similarities, making the concept non-trivial.

For property-based accounts of similarity, the too many similarities problem is a consequence of the Ugly Duckling Theorem. The Ugly Duckling Theorem says that if properties are extensionally defined and closed under complementation and negation (i.e. form a Boolean lattice), then any two objects share the exact same number of properties. This means that any two objects are exactly as similar as any other two objects. To see this, consider a finite universe of N objects. There are 2^N ways to group them into sets, each of which corresponds to a property. Code each set with an N digit binary string where each digit stands for the membership, 1, or not, 0, of the respective object in that set; there will be 2^N such strings. Pick two objects arbitrarily and order the strings lexicographically with respect to their digits. On the first quarter of the strings, they will agree on membership, i.e. the strings will begin 11 . . . , on the next two quarters they will disagree (strings 10 . . . and 01 . . .), but on the final quarter they will agree again (strings beginning 00 . . .). So the two objects agree on exactly 1/2 of their properties, i.e. on 2^{N-1} properties. Since they were chosen arbitrarily, the result holds for any two objects.

The Ugly Duckling Theorem is near trivial, but it makes a crucial point. If you want to say two objects are similar because they share properties, you need to privilege some properties over others. If you accept all boolean combinations of properties as themselves properties of equal status, your claim is trivial. This result was proved (and named) in Watanabe (1969), but had surely been known for a long time before that. The first edition of Goodman (1976) alludes to it and Goodman (1970) gives the argument explicitly, without any evidence of influence from Watanabe.

For isomorphism-based accounts of similarity, the Newman Problem plays the same role as the Ugly Duckling Theorem. In response to Russell's early structural realism, Newman (1928) pointed out that an isomorphism exists between any two sets of the same cardinality. If we demand that the isomorphism preserve relations found in one of the structures, it will always be mathematically possible to define relations in the second structure which satisfy it, so long as the cardinality of the two structures is the same. McLendon (1955) pointed out that the problem is even worse. Since any set of objects can be subdivided into arbitrarily many subsets, an isomorphism exists between structures defined over any two sets regardless of cardinality. Since some such isomorphism exists between any two structures, the naïve claim that two structures are similar because an isomorphism exists between them is trivially true of any two structures.

Again, the moral is that some source of bias is needed. We must privilege some isomorphisms over others, or some relations over others, in order for the claim that similarity occurs whenever two objects are isomorphic to be non-trivial. In the account of similarity developed here, the properties and relations of the scientific view of the world provide the source of our bias. In order to see how this will work, let's first look at some examples.

2.2 Examples of Similarity in Nature

Aristotle's signet ring provides a good starting point for thinking about similarity in nature. Of course, the ring wearer intends the wax to take the shape of the ring, but that it does so does not depend on his wishes, but on the properties of the ring and the wax. Examples of similarity of shape such as this can be found everywhere in nature. For instance, the relationship between each of these two objects might plausibly be considered one of similarity:

1. a footprint and the foot which made it
2. a shadow of a tree and the tree which cast it
3. a fossil of a mollusc and the original mollusc

In each case, there are two shapes, and some gross similarity between the shapes in virtue of the fact that one caused the other. In the case of the footprint and the fossil, the causal relationship involves direct contact, or at least spatial coincidence. In the case of the shadow, the fact that light travels in straight lines ensures that rays from a single source will induce a projection of any three dimensional object onto the surface behind it.

Of course, direct contact is not a necessary condition for sameness of shape. Two shapes may be similar in virtue of each having separate contact with the same object, for instance

4. two footprints, both made by the same foot

More complex causal relations can also induce similarities of shape. Consider for example

5. the pattern of colors on a butterfly wing and that on the wing of its offspring

In this case, there is a similarity of shape (and color), but the causal process involves the transmission of genetic material, epigenetics, sameness of environment and diet, and many additional factors. Furthermore, it is not just shape which is similar in this case, but also behavior and life cycle.

In each of these examples, we might argue that the two similar objects share the same properties, and that their similarity is reflected in the fact that these properties take the same value for each object. For instance, the foot is 11 inches long, and its footprint is also 11 inches long; or, the butterfly's wing is yellow and its offspring's wing is also yellow. But similarity may also hold purely at the structural level; consider the relationship between

6. the (relative) width of a tree ring and the (relative) amount of rain during the year in which it was formed
7. fluctuations in the ratio of black to white peppered moths and fluctuations in the ratio of dark to light colored tree bark in Sherwood Forest

In 6, the properties at issue are different: widths and volumes are different sorts of properties. The similarity relation holds not between the sets of properties, but between the pattern of instances of one property and the pattern of instances of another. This is the more general type of similarity which can be found in nature, and instances in which the properties happen to be the same are just special cases of it.

2.3 Objective Similarity Defined

Nature is filled with structures. An animal, a population, a spatio-temporal region: each of these has internal structure, i.e. a pattern of parts with properties standing in various relations to each other. The same natural structure may be characterized in many different ways—in the present project, we allow ourselves any method of doing so which can be found in the scientific description of the world, and deny any which is artificially or arbitrarily constructed. Science then becomes the source of bias which privileges some relations and mappings over others, avoiding the Newman Problem.

Formally, we can represent structures as set theoretical objects, e.g. $\langle \mathcal{X}, R_1, R_2, R_3, \dots \rangle$, where \mathcal{X} is the set of parts and R_1, R_2, R_3, \dots are n -ary relations which hold between them (a property is just a unary relation). We say a structure is *trivial* if all R_1, R_2, R_3, \dots are either empty or equivalence relations over all \mathcal{X} . The natural similarity relationship between two set theoretical objects is a *homomorphism*, or structure preserving mapping. If $\langle \mathcal{X}, R \rangle$ and $\langle \mathcal{Y}, S \rangle$ are two structures with R and S binary relations, then a map $\phi : \mathcal{X} \rightarrow \mathcal{Y}$ is a homomorphism just in case xRx' if and only if $\phi(x)S\phi(x')$ for all $x, x' \in \mathcal{X}$.

Of particular interest are structural relationships defined by orderings. Consider, for example, the case of tree rings and rain. Take $\langle \mathcal{R}, \leq_C, \leq_W \rangle$ to be the set of tree rings with two orderings defined on them. For $r, r' \in \mathcal{R}$, $r \leq_C r'$ iff r' is at least as far from the center of the tree as r , and $r \leq_W r'$ iff r' is at least as wide as r . Take $\langle \mathcal{W}, \preceq_T, \preceq_V \rangle$ to be the collection of annual rainfalls in the forest during the life of the tree. For $w, w' \in \mathcal{W}$, $w \preceq_T w'$ iff w' occurred at least as late as w and $w \preceq_V w'$ iff w' was at least as great in volume as w . There exists a map $\psi : \mathcal{W} \rightarrow \mathcal{R}$ such that $w \preceq_T w'$ iff $\psi(w) \leq_C \psi(w')$ and $w \preceq_V w'$ iff $\psi(w) \leq_W \psi(w')$. Furthermore, it is in virtue of the inverse of this map, ψ^{-1} , that dendrochronologists are able to learn about past rainfall by examining the rings of old trees.

The tree rings and rainfall, just like a foot and its footprint, are similar as the result of a direct causal relationship. We might plausibly claim that the quantity of rainfall in a given year caused the width of the respective ring, or the shape of the foot caused the shape of the footprint. Let us say that a homomorphism which exists as a consequence of a causal process is *induced* by that process.

Homomorphisms which are induced by causal processes establish similarity relations, but they do not exhaust the similarities which can be found in nature. Consider two examples. First, the relationship between a foot and a plaster cast made from one of its footprints. The shape of the footprint is induced by its causal relationship to the foot, while the shape of the cast is induced by its causal relationship to the footprint. It doesn't matter whether these two causal relations are in any way related (in this case, plausibly, they are not)—what matters for ensuring similarity is that the structure preserved by the second causal process is the same as the structure preserved by the first causal process.

Second, consider the relationship between two footprints made by the same foot. In this case, there exists a structure preserving homomorphism between the two footprints, but it is not induced directly by a causal process. Rather, it is the fact that both footprints stand in the same causal relationship to a third structure, the foot, which ensures they are similar.

These considerations motivate the definition of a *similarity map*.

Similarity Map – call a homomorphism ϕ from structure A into structure B a *similarity map* if any one of the following hold

- ϕ was induced by a causal process in which the structure of A causally influenced the structure of B .

- $A = B$.
- There exist maps ψ and ψ' and structure C such that ψ is a similarity map from A into C , ψ' is a similarity map from C into B , and $\phi = \psi' \circ \psi$.
- There exist maps ψ and ψ' and structure C such that ψ is a similarity map from C into A and ψ' is a similarity map from C into B such that $\psi' = \phi \circ \psi$.

The first condition establishes the base case. The second condition ensures reflexivity. The third condition is motivated by the example of the cast and the foot. Since the definition is inductive in structure, it allows for, e.g. footprints made by the plaster cast to be similar to the original foot, or the pattern on a butterfly's wing to be similar to that on the wing of its great-great-grandmother. The fourth condition is motivated by examples like footprints made by the same foot, or widths of tree rings in the same forest.

Is the set of similarity maps symmetric, i.e. if θ is a similarity map from B into A , is θ^{-1} a similarity map? Symmetry follows from clauses 2 and 4. To see this, consider the instance of clause 4 for which $C = B$. By stipulation the similarity map $\psi (= \theta)$ exists. ψ' exists and is a similarity map by clause 2 and the stipulation that $C = B$. Then $\phi (= \theta^{-1})$ is a similarity map as it satisfies the constraint that $\psi' = \phi \circ \psi$ since $\psi' = \theta^{-1} \circ \theta$.

Finally, notice that the similarity which holds between two structures in nature is established by a *partial* map between them. Rings are not all there is to a tree; shape of the sole is not all there is to a foot; a shadow does not preserve the full three dimensional structure of the object which cast it. These considerations motivate the definition of *objective similarity*:

Objective Similarity – Two natural structures A and B are *objectively similar* if there exist non-trivial substructures $A' \subseteq A$ and $B' \subseteq B$ such that

1. There exists a homomorphism $\phi : A' \rightarrow B'$
2. The homomorphism ϕ is a similarity map.

Note that this definition applies to all the above examples of similarity in nature. For example, the relevant structure preserved in cases of similar shape is relative distance between points, the causal process is direct contact (the footprint), obstruction (the shadow), or genetic transmission (the butterfly).

The insistence that the substructures be non-trivial rules out trivial or vacuous similarities where there is a causal relationship but no structure preserving map. A cocoon is not similar to a moth, a hole in the ground is not similar to a stick of dynamite, and a mosquito bite is not similar to a mosquito. In each case, there is a causal relationship, but no shared structural relations between cause and effect, so only trivial homomorphisms exist between the two structures.

By allowing ourselves to assume that there are natural (i.e. scientific) categories and causal relations in the world, we've defined a notion of similarity which is objective, in the sense that it does not depend upon the judgments of any particular agent. We'll see in Section 4 that this is the exact notion of similarity employed by many psychologists in analyzing mental representations. Before we get there, however, we need to discuss the general relationship between objective similarity and representation in nature.

3 From Similarity to Representation

The attempt to define representation in terms of similarity faces a fundamental problem: similarity is symmetric, but representation is not. After discussing the classic presentation of this problem in Goodman, we'll define a type of representation which avoids it, N-representation, or representation in nature. N-representation avoids Goodman's worry by making use of the asymmetry in the causal relations we've already allowed ourselves. Happily, then, we can develop a notion of representation from objective similarity which requires no additional assumptions.

But is N-representation really representation? Van Fraassen thinks not. He argues that the considerations raised by Goodman demonstrate that not only directionality, but human purpose is required for representation. I conclude the section by arguing that van Fraassen begs the question. By directing attention to examples from science rather than art, I demonstrate that N-representation shares qualitative features with intuitive examples of representation.

3.1 The Challenge: *Directionality*

Nelson Goodman's *Languages of Art* begins with a sustained argument that resemblance is inadequate as a basis for representation. His fundamental point is that representation and symmetry have different logical properties.

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, p. 4)

Goodman argues that a) similarity is reflexive, but representation is not; b) similarity is symmetric, but representation is not; and c) obvious examples of similarity are not instances of representation.

The properties of similarity Goodman identifies hold for objective similarity. Objective similarity is reflexive due to the second clause of the definition of similarity map. The symmetry of objective similarity follows from clauses 2 and 4, as discussed in Section 2.3. The examples of twin brothers and cars from the same assembly line both satisfy the definition of objective similarity. But Goodman's critique applies to a bare notion of similarity, one which simply has the logical properties he identifies. Our notion of objective similarity is much richer, with a built in source of directionality waiting to be exploited.

3.2 N-Representation and Its Properties

Causality is directional. In our definition of similarity map, we explicitly introduced reflexivity and symmetry in order to derive a relation with the properties of similarity from causality. In order to find those structures which are both similar and such that one represents the other, we simply need to reintroduce that directionality. The fundamental idea here is that one structure can represent another only if it is causally downstream from it.

Causally Downstream – Given two natural structures A and B , A is *causally downstream* from B whenever $A \neq B$ and there exist structures C_1, C_2, \dots, C_n such that B causally influenced C_1 , C_n causally influenced A , and, for each $1 \leq x \leq n - 1$, C_x causally influenced C_{x+1} .

Combining this directionality with the requirements of objective similarity we can define representation in nature.

N-representation – Given two non-trivial natural structures A and B , A *N-represents* B whenever both

1. A is objectively similar to B .
2. A is causally downstream from B .

Of course, we could have defined N-representation directly, without detouring through a general notion of objective similarity. But in order to do so, we would have had to build the notion of structural similarity into the definition nevertheless.

The first point to notice is that N-representation has the appropriate logical properties to count as a representational relationship. It is neither reflexive nor symmetric; it is not satisfied by the examples of similarity without representation provided by Goodman. For a more thorough argument that causally induced homomorphisms satisfy the basic logical requirements for a theory of representation, see Bartels (2006).

In what sense is N-representation “representation”? In the very intuitive sense of the re-presentation of information. A relevant litmus test here is whether A can be used to learn about B . This was our initial explanatory target—how can a mental representation be used to learn about the world? If A re-presents information initially presented in B , then A can be used to learn about B . The insistence on objective similarity ensures that A N-represents B implies that A contains information about B , i.e. can be used to learn about B . The restriction to causal direction captures the asymmetrical nature of information flow. In particular, if A represents B and B represents C , then B contains at least as much information about C as A does. It is impossible to learn more about C from A than from B ; increasing causal distance from a structure means less information about that structure.

Consider, for example, the amount of information a forensic podiatrist can learn about a suspect’s boot from a footprint. A footprint in damp, thick mud may leave lots of detail, the exact placing of tread marks, irregularities in the wear of the sole, etc. Conversely, a footprint in very wet mud with low viscosity may merely preserve the rough outline of the edge of the boot. The footprint in thick mud is *more similar* to the boot than the footprint in wet mud precisely because it contains more information about the boot. At a formal level, a larger substructure of the boot satisfies the objective similarity relation with respect to the footprint in thick mud than with respect to the footprint in wet mud.

Now suppose the forensic podiatrist takes a plaster cast of the footprint. He may be able to learn about the boot *more easily* from this cast than from the initial footprint (it’s easier to handle, to measure, to transport; it won’t disappear with a heavy rain). But the cast can’t contain more information about the boot than the initial footprint—the process of making the cast can’t “put in” any new information, it can only extract information that’s already there. Consequently, the footprint itself will always be more similar to the boot than the cast.

This is just a specific instance of a general feature of information transfer familiar to anyone who has copied a VHS tape. A copy of a VHS tape always looks worse than the original, more fuzziness and noise, and a copy of a copy even more so. The same holds for photocopies: a copy of a document always has more noise than the original document. *Information degrades as it is transferred.* This is a general feature of the re-presentation of information captured by the notion of N-representation.²

3.3 Directionality Revisited

N-representation is a relation which exists in nature, independent of any human intentionality, agency, or information use. It is a representational relation in the sense that a) it satisfies the logical constraints on a such a relation (in particular, it is directed), and b) it captures the intuitive constraint that *A* represents *B* implies that *A* can be used to learn about *B*. However, some have argued that mere directionality and informativeness are not enough to constitute “representation.” In particular, van Fraassen explicitly considers a relation very like N-representation, but denies that it constitutes a species of representation:

On the macroscopic level too we can think of processes that connect two situations separated in time or space. These could be so correlated that it would be possible in principle to get information about the one by inspecting the other—provided of course we knew of that correlation! But that something *could be done* does not mean that *it is done*. That something could be assigned a representational role does not mean that it has one. (van Fraassen, 2008, p. 156)

For van Fraassen there is “no room for ‘representation in nature’, in the sense of ‘naturally produced’ representations that have nothing to do with conscious or cognitive activity or communication” (van Fraassen, 2008, p. 24).

If van Fraassen is right, then N-representation is not representation. Of course, van Fraassen better not be right, otherwise there would be no hope for a non-circular analysis of mental representation. If conscious activity is required to explain any instance of representation, then mental structures can only constitute representations if they are taken as such by a conscious agent. But the very concept of a conscious agent already assumes representational mental structure. *Regress.* Where has van Fraassen gone wrong?

Van Fraassen (2008) begins with an elaboration of the considerations of Goodman (1976). Whereas we have focussed on the bare logical structure of representation as emphasized by Goodman, namely its directionality, van Fraassen emphasizes the source of this directionality in Goodman’s examples, namely human intent. For van Fraassen, representations are inherently *perspectival*, the source of this perspective is the *choice, goal, or intent* of a particular representation user, and the mark of this perspectival quality is the fact that representations are *distorted* with respect to their target.

For example, consider the ancient competition between Alcamenes and Phidias (reported in Gombrich, 1960, pp. 191–2). The two sculptors are competing to produce a statue of Minerva to be placed upon a high pillar. Alcamenes produces a sculpture of a beautiful woman, while Phidias produces a sculpture with stretched and enlarged features of the face. Yet, when the sculptures are placed upon the high

²Pace Dretske’s unfortunately named “Xerox principle” (Dretske, 1981, p. 58), which denies this basic empirical feature of information transfer.

pedestal, that of Phidias is much more pleasing to the eye when viewed from below. Distortion in the sculpture made it a better representation when viewed from the relevant perspective.

Consider another example: a caricature of Richard Nixon. Such a caricature is more effective, succeeds better at representing Nixon, when it emphasizes certain features, say the length of the nose, the beadiness of the eyes, and the slackness of the jowls. Greater similarity with Nixon's actual features does not equate with better representation of Nixon in a political cartoon. Van Fraassen concludes from examples such as these that distortion, more importantly *the choice of which distortions to employ*, is essential for representation.

It seems then that distortion, infidelity, lack of resemblance in some respect, may in general be crucial to the success of a representation. This does not rule out that resemblance in some other respect may be required. Yet even when that is the case—and it may be a special case—the *choice* of those respects in which resemblance or a specific kind of distortion is required, and those for which just anything at all will do, will have to be seen as crucial as well. (van Fraassen, 2008, p. 13, emphasis in original)

For the statue of Minerva, distortions contribute to better representation because the representation is indexed by a perspective, i.e. the angle of view. The caricature also is indexed by a particular perspective, though in this case not a literal viewpoint, but a stance towards Nixon as buffoonish and, perhaps, “tricky.”

It is critical to van Fraassen's argument that perspective is a matter of choice. Furthermore, this choice cannot be found within the object itself—the representational properties of an object may change with its use. Van Fraassen illustrates this point by considering a famous photograph of the Eiffel Tower by Doisneau.

If I send it to you from Paris as a postcard, with the single note “Wish you were here!”, then it is itself a photo of the Eiffel Tower. If I insert it into the book I am writing about photography, then it represents the famous photo by Doisneau . . . In other words, if it is an image of something at all then *what it is an image of* depends on the use, on *what I use it to represent*. (van Fraassen, 2008, p. 21)

These considerations drive van Fraassen to state his “*hauptsatz*,” or “fundamental theorem,” of representation:

There is no representation except in the sense that some things are used, made, or taken, to represent some things as thus or so. (p. 23)

It is this *hauptsatz* which ensures that there can be no natural representation; if we accept it, then we must acknowledge that N-representation is not a species of representation. But notice that there are two distinct steps leading to van Fraassen's *hauptsatz*: first, the observation that representations are inherently perspectival; second, the claim that the setting of perspective is a matter of choice or use. I wish to accept the first step, but deny the second. If we consider the kinds of examples Goodman considered *in the context of artistic representation*, e.g. caricatures, then van Fraassen's second step seems natural. It looks much less so, however, if we consider other examples of representation more closely.

Take, for instance, the example of the Mercator projection. This is a technique for making a two dimensional map of a spherical surface—the map of the world hanging in your second grade classroom was likely the result of a Mercator projection. One can make a Mercator projection very simply if one has a transparent globe with the features to be mapped outlined on it in black. Wrap a piece of paper around the globe along one of its great circles and shine a light from inside the globe. The shadow cast on the wrapped piece of paper is the Mercator projection and can be traced. To make the map in your second grade classroom, the paper was wrapped around the globe along the equator.

A Mercator projection is inherently perspectival, and it distorts the true pattern of lines on the surface of the globe. The choice of which great circle around which to wrap the paper sets a perspective. Features on the surface of the globe are more and more distorted in the projection as they get further from the point of contact with the paper. At the limits, e.g. the poles on a typical Mercator map, distortion becomes infinite. In particular, the poles, which are mere points on the surface of the globe, are stretched to lines on the map which fill the entire top and bottom edges of the paper. The choice of perspective is related to this distortion. In particular, we choose the equator, or some very close meridian, in order to minimally distort those landmasses near the equator on which the preponderance of the human population of the earth lives. So, it is indeed our interest in the world, the use to which we intend to put the map, which determines the choice of perspective.

Notice, however, a crucial difference between the Mercator projection and the caricature. In the caricature of Nixon, there are many choices to make: how long the nose should be, how slack the jowls, how beady the eyes. There is no strong relationship between these choices; the length of the nose does not in any way determine the slackness of the jowls. In the Mercator projection, there is a single choice to make: which great circle around which to wrap the paper. Once this single choice is made, the distortions of the globe's surface in the map are entirely *systematic*. In the case of the Mercator projection, there is one degree of freedom, one choice to be made, in the case of the caricature, there are many (arguably infinite) degrees of freedom.

But N-representation is systematic in precisely this sense! This systematicity is a result of the requirement that an N-representation be induced by a causal process. A foot may step in mud at many different angles. Once the foot does step, however, the relationship between it and the footprint is completely determined by the angle and the properties of the mud. The relationship between the width of the heel and the width of the back of the footprint and that between the width of the toes and the width of the front of the footprint cannot be varied independently. Likewise, once the position of the sun is set, the pattern of the tree's shadow is uniquely determined. In these examples, there *is* a perspective needed to establish N-representation, but that perspective is determined by a natural process, not some agent's conscious choice.

Van Fraassen's argument begs the question against N-representation. He begins by considering representation in art; he observes the role of choices; and he derives his *hauptsatz*. If instead we begin by looking at systematic examples of representation, such as the Mercator projection, the role of choice in establishing the perspectival quality of representation seems much less essential. Once we observe that shadows and footprints share essential features with the Mercator projection, it is natural to subsume them all under a common notion of representation.

Of course, this is not to deny that van Fraassen's considerations apply to some representations. We might acknowledge that there is some sense in which the photo in the art book represents a famous

photograph by Doisneau, and not the Eiffel Tower. But there is also something about the photo which makes it suitable to be used as a postcard. This something is captured by the fact that the photo N-represents the Eiffel Tower. N-representation can thus explain how an object is suitable for use as an intentional representation. There is room in this world for more than one type of representation, and N-representation can help us to better understand more elaborate examples of representation, whether we find them in art, science, or the mind.

4 Mental Representation

Now that we've developed a philosophical account of similarity in nature and its role in supporting representation, we can return to psychological practice and try to make some philosophical sense of it. In particular, we'll survey a number of psychologists who take mental representations to be N-representations which guide behavior. By looking at a sequence of increasingly nuanced views, we'll tease out the exact role of similarity in this account of representation. After a brief digression on the consequences of this view for the debate between symbolic and imagistic theories of representation, we'll conclude by looking at the problem of misrepresentation. The view developed here shifts the burden for explaining misrepresentation onto an analysis of the goals and actions of the organism. Nevertheless, the theory of N-representation can give some insight into the differential role which more or less informative mental representations play in generating more or less appropriate actions toward a goal.

4.1 "Functioning" N-Representations

The psychologists surveyed here identify two marks of a mental representation: first, that it constitutes an N-representation; second, that it be *functioning*.³ It is important to distinguish the use of the term "function" here from its use in the philosophy literature (e.g. "functionalism" or "functional role semantics"). Let's look at some representative passages in order to understand the notion of "functioning" relevant for this theory of mental representation.

Stephen Palmer's views are representative of those which can be found throughout the study of perception. He identifies representation with a causally induced homomorphism, i.e. an N-representation:

... a representational system can be analyzed as a homomorphism: a mapping from objects in one domain (the external world) to objects in another domain (the internal representation) such that relations among objects in the external world are mirrored by corresponding relations among corresponding objects in the representation. ...

The causal factor underlying the homomorphism is important for two reasons. One is that for the representation to be current, as a perceptual representation must be, it requires constant updating. A causal chain from events in the external world to events in the internal representation is an ideal way (though not the only way) to achieve this. The other is that for the representation to be authentic, rather than accidental, there must be some linkage to

³Structurally, this view bears some similarity to so-called "two factor" versions of the functional role theory of mental representations (for a survey, see Perlman, 1997), though the details differ substantially from any pre-existing two factor view of which I am aware. In particular, it would be a mistake to conflate "functional role" in the sense used here with "conceptual role." The closest philosophical analog to this view is in the work of Robert Cummins, discussed later in this section.

the world it represents. Again, a causal connection seems to be the ideal solution. (Palmer, 1999, p. 77–8)

One reason for the popularity of this view in the study of perception is the overwhelming evidence for the importance of causally induced structural homomorphisms in the physiology of perception. Cones in the eye, for example, are arrayed across the retina like the photoreceptors in a camera. The spatial pattern of excitation on the retina is projected through the lateral geniculate nucleus, onto the primary visual cortex at the back of the brain, then forward through the later areas of visual processing in a sequence of structure preserving maps. At each stage of this processing chain, the pattern of neural activation constitutes an N-representation of the stimulus.

The N-representations in a perceptual system are not full-blown representations according to Palmer unless they are exploited for reasoning about the world.

Causally driven structural similarity between a model and its object may be a necessary condition for something to function as a representation, but it is not sufficient. A further requirement is that there be processes that use it as a representation in the sense of taking it as a surrogate of the world to which it corresponds. (Palmer, 1999, p. 78)

The mark of a functioning representation is not to be found in the intent or choice of an agent, but in the actions the system performs. Palmer uses the example of a thermostat here, and the relation between its internal bimetallic strip and ambient temperature. The position of the strip's end N-represents changes in ambient temperature. The crucial additional factor here is not what the thermostat was built *for* (regulating temperature), but what it *does*, namely respond differentially to changes in the position of the strip's end. It uses the position of the strip's end as a surrogate for the world in the sense that it responds differentially to the position of the strip, not to changes in temperature directly.

N-representations again emerge as a natural account of representation in the study of action and navigation. The abilities of bees, wasps, mice, etc. to navigate novel paths through familiar environments is easily modeled on the assumption of a structured internal representation, standing in a causally-induced homomorphism toward the physical environment. The fact that these paths are *novel* demonstrates they result from the use of representations rather than mere habit. The fact that these paths depend upon the *familiarity* of the environment demonstrates that they are causally induced by the organism's exploration of that environment. An example here is the model of Deutsch (1960), which posits that familiar environments are represented by a cognitive network of nodes (standing for landmarks) with connections signifying topological connectedness. Such a model can account for empirical findings about the navigational behavior of wasps, bees, mice, etc. (see Gallistel, 1980, Ch. 11, for a detailed discussion).

Like Palmer, Gallistel requires more than mere N-representation, he requires that N-representations be *functioning*, or “put to use”:

The exploitation of the correspondence to solve problems in the one domain using operations belonging to the other establishes a functioning isomorphism: an isomorphism in which the capacity of one system to represent another is put to use. (Gallistel, 1993, pp. 15–6)

Gallistel's notion of functioning requires that an organism's differential responses to an N-representation be systematic:

Isomorphisms exist between systems, not between sets of entities. . . . The representing system must perform operations on the entities generated by the mapping, and there must be a correspondence between the operations the representing system performs with the representatives and the relations or processes in which the represented entities play roles—for example, the correspondence between numerical addition and combining two weights on a pan of a balance. (Gallistel, 1993, pp. 24)

In technical terms, Gallistel requires that the homomorphism between a mental representation and the world preserve not only relations (static structure), but also *operations* (dynamic structure). Changes in the mental representation must systematically correspond to changes in the world. For instance in the Deutsch model, changing activation at nodes in the wasp’s internal representation should correspond to changing positions within its environment.

One way to make sense of Gallistel’s notion of system is in terms of sets of N-representations. Consider a set of N-representations such that they N-represent the environment via the same causal mechanism, and small differences between the N-representations correspond to small differences in the environment. Now suppose a second, internal causal factor (an operation) influences which N-representation is instantiated, and that application of this internal operation corresponds to behavior by the organism which brings about the corresponding state of the environment. As activation spreads through the wasp’s cognitive map, it moves through its environment, changing which landmark is in view.

Mental representations N-represent the environment, and consequently they are objectively similar to the environment. What role does this similarity between representation and environment play in subserving the systematic use of mental representations described by Gallistel? A nuanced answer to this question can be found in the work of Roger Shepard.

Shepard developed his view while performing experiments on spatial reasoning. He discovered, for example, that the time it takes to recognize a shape varies linearly with the angle of rotation of the stimulus shape from its initial presentation. Moreover, response times can be manipulated by asking subjects to visualize rotation of the shape ahead of time (Shepard, 1975). Subjects can recognize details of spatial structure with an enormous degree of precision, e.g. which angles of intersection between two lines are possible views of the corner of a cube and which are not. They perceive continuous rotations or transformations when presented with static stimuli in quick succession (as when two lights a small distance apart flashing in sequence are perceived as a single light moving rapidly back and forth) (Shepard, 1981). Shepard developed a unified framework for modeling these diverse phenomena.

Shepard argues that systems of the sort described by Gallistel implement *second-order* isomorphisms:

[W]hat enables a given internal process to represent or to stand for a particular external object is the way that that internal process functions within the organism—particularly in his relation to that object and to other similar objects. The essential things are, in particular, (a) that there be an orderly causal connection between the external objects and the internal process that is to represent it, and (b) that internal representations corresponding to similar external objects be functionally related so that a response learned to one will tend to generalize to another (as required by second-order isomorphism). When, in addition, there is a degree of first-order isomorphism, it is not there to enable the internal representation to refer to its external object. It is there, rather, to provide the mechanism for the realization of the

necessary second-order isomorphism and, especially, to provide the structural information needed for more differentiated responses of selective attention, analysis, transformation, and manipulation. (Shepard, 1975, p. 92)

Let's unpack this into our terminology. By Shepard's lights, the most significant feature of a mental representation is not that it constitute an N-representation of the environment in the static ("first-order") sense. More important is that sets of like objects generate the same, or closely related, responses. Insofar as the set of like objects in the environment and the set of generated responses themselves constitute legitimate structures for scientific study, the relationship between these two sets will also satisfy the definition of objective similarity. Nevertheless, the relationship between these structures is importantly different from the static structural similarities we have examined so far, it is dynamic and hypothetical. The objective similarity here is between *possible* stimuli and *possible* responses, not a particular given stimulus and the particular response it generates.

Consider again the example of the wasp. The wasp is familiar with its environment and has generated an internal map which N-represents that environment. By Shepard's lights, the mark of full representation here is that like inputs (say, a landmark viewed from different angles) generate like responses (say, crawling in the direction of its nest). It is this second-order map between inputs and behaviors which demonstrates representation. The first-order map between cognitive structure and environment is important only insofar as it subserves the second-order map. It is "the mechanism for the realization of the necessary second-order isomorphism."

Shepard (1981) resists calling the second-order map which marks representation a similarity, preferring instead the term "complementarity." His guiding analogy for thinking about complementarity is the relationship between a key and a lock.⁴ This analogy nicely moves beyond Aristotle's signet ring example by adding a dynamical, functional component. Provided the dynamical structures which stand in this second-order relationship continue to satisfy the requirement that they appear in the scientific view of nature, they will still satisfy the definition of objective similarity. Thus a dynamic similarity obtains between possible objects the organism might encounter and possible behaviors it might generate in response.

Shepard speculates that first-order objective similarity *may* subserve second-order functioning complementarities between behavior and environment. The view that first-order similarity is actually *necessary* for the successful representation of and response to nature has been defended at length by Robert Cummins (e.g. Cummins, 1996; Cummins and Poirer, 2004). Cummins takes the mark of a representation to be the potential it offers for reasoning about the represented. In order to make sense of how an organism comes to exploit a representational structure, its representational content must be present antecedent to the organism's use of it. This is what Cummins has called "unexploited content—content a representation *has*, but which the system that harbors it is currently unable to exploit" (Cummins et al., 2006, p. 195). Cummins argues that this unexploited content must be present *first* before the representational structure can come to be used, or exploited, by some function. His solution for explaining unexploited content is to identify representations with structures isomorphic to a target. Then the internal structure of the representation represents the structure of the target in virtue of this isomorphism, independent of the functional role the representing structure plays within the organism.

⁴c.f. Cummins, 1996, p. 102.

Despite many commonalities between the view of Cummins and that presented here, we diverge on the preconditions for similarity. Cummins identifies representation with the mere existence of an isomorphism between representation and represented, and does not restrict attention to causally induced mappings as we have done. He argues that the Newman Problem (too many isomorphisms) need not be solved in order for isomorphic representations to play the relevant explanatory role in understanding behavior. It is the fit between a representation and the target of an organism’s actions which explains how a representation successfully subserves action; if the representation fits additional, irrelevant targets as well, this fact in no way undermines this explanatory role. Consequently, “[t]he fact that [the picture theory of representation] makes representational contents non-unique, . . . presents no problem, provided it does not make representational content indiscriminate” (Cummins, 1996, p. 102).

I’m afraid that I must agree with Cummins’ conclusion, while disagreeing with its consequences for his view. To take the Newman Problem seriously is to realize that the isomorphism view of representational content *is* indiscriminate without a prior constraint on permissible relations and mappings. *Some* mapping exists between *any* two structures. Here, we’ve insisted that homomorphisms be causally induced in order to constrain the number of permissible mappings. Our account of representation does indeed leave contents non-unique, but no longer indiscriminate. Taking causality as the relevant source of bias has the further happy consequence of explaining *why* the unexploited content in a mental representation represents, not just *how*. A mental representation represents the world *via* its similarity with the world and *because* it was causally induced by the world.

A common theory of mental representation has emerged from the psychologists surveyed here. Mental representations represent the world in virtue of their objective similarity to the world. This similarity exists and is relevant because it is causally induced. The most important feature of this similarity is that it be dynamic, i.e. that mental operations on an internal representation systematically correspond to physical operations on the represented target in the world. This second-order similarity is most likely (arguably, necessarily) subserved by a first-order similarity between static representations and static structures in the world.

A final question concerns the locus of misrepresentation in the view developed here. Before we tackle that problem, however, a brief digression on the relationship between the requirement of homomorphism and the debate between symbolic and imagistic theories of representation.

4.2 Brief Digression: *Symbols and Similarity*

In the twentieth century, the rejection of similarity went hand in hand with the rise of symbolic theories of representation. Likewise, philosophers known for championing the view that mental representations represent in virtue of their homomorphic relationship to structures in the world are often known also for embracing alternatives to the symbolic approach, such as connectionism (e.g. Paul Churchland, 2002, 2007). Furthermore, several of the psychologists discussed here (e.g. Shepard and Kosslyn) have defended the importance of homomorphisms in the context of arguing for an imagistic theory of representation. Nevertheless, it is important to emphasize that the requirement of objective similarity can be met *even by symbolic systems*, and thus the considerations raised here cut across the symbolic vs. imagistic debate.

An easy example to see this is that of navigation in a familiar environment discussed above. I claim, following Deutsch and Gallistel, that bees, wasps, etc. navigate a familiar environment using a representation of that environment. This representation is an N-representation, and it is in virtue of the

objective similarity between this representation and the environment that the organism is able to navigate successfully. Now, it so happens that Duetsch and Gallistel conceive of this representation as a network of nodes, where each node stands for some landmark in the environment. But could we achieve the same representational goal, preserving the requirement of objective similarity, with a symbolic representation?

Not only *can* this goal be achieved, but indeed it *must* be achieved if the symbolic system is to be of any use for navigational purposes. Such a symbolic system might use arbitrary symbols a, b, c , etc. to represent landmarks and some relational symbol R to represent the “next to” relation. Then, the symbolic representation of the environment would consist of a list aRc, cRd, dRa , etc. of sentences in this symbolic system. If it is faithful to the environment, this symbolic representation will stand in a homomorphic relationship to it, and if it was learned, this homomorphism will be causally induced. So, while the relationship between each symbol and each landmark will be arbitrary, the relationship between the symbolic system *as a whole* and the environment will constitute an N-representation, and it is in virtue of the objective similarity between the symbolic system and the environment that it can be used to navigate successfully.⁵

Now, finally, we can resolve the puzzle of our first paragraph. Why is it that a pencil which I have stipulated represents Canada cannot be used to reason about Canada? Because it is not part of a structure. Suppose I stipulate this pen represents the United States, and this eraser represents Mexico. Still, nothing can be learned about Canada, the U.S., or Mexico. Once I add structure, however, say by organizing the pencil, pen, and eraser in a line such that each is to the left of the next, the structure as a whole can be used for reasoning. If, for instance, I discover the homomorphism between this structure and North America (which maps “to the left of” onto “to the North of”), I can use the structure to behave in the world. For example, if I want to get to Canada from the U.S., I can use this representation to discover that I should go North.

In order for reasoning to work here, it doesn’t matter whether the relation between primitive parts of the structure is spatial (as when I arrange them on a desk) or symbolic (as when I write a list of sentences in which a relation symbol stands for “to the North of”). What matters for successful reasoning is that the representation have, or be part of, a structure, and that this structure stand in the correct relationship to the world, namely that it N-represents the world. Then the objective similarity between representation and world can be exploited to guide action.

4.3 Misrepresentation Demystified

Some philosophers take the possibility of misrepresentation as constitutive of the concept of representation. Yet the possibility of mis-X presumes a normative standard for evaluating X, whatever X might be. This worry motivates the move to teleosemantics in, for example, the work of Dretske (1988, 1995). What is the source of the potential for misrepresentation on the view developed here? If N-representations are merely a consequence of causal relations, how can they possibly misrepresent?

Once an N-representation is exploited by an organism, then it can be used successfully or unsuccessfully. So, requirement that a mental representation constitute a functioning N-representation permits at least one source of potential representational error, namely error as a mismatch between the goal

⁵Of course, the use of a network rather than a symbolic system is crucial in the Deutsch and Gallistel model because the actions of the organism are explained in terms of spreading activation along a gradient through the network. The example considered here addresses merely the requirements on representation, not the mechanism for using representations to guide action.

of the organism in using the representation and the actions it generates from that representation. Of course, there are still puzzles about normativity here: what determines the organism’s goals; whence its agency? However, these are not puzzles about the normativity of content. Misrepresentation of this kind reduces entirely to misuse, and the onus for explaining representational error falls not on the theory of representation, but on the theory of action.

This is the strategy pursued by Cummins for explaining representational error. He argues that theories which cash representation out purely in terms of functional role cannot distinguish between content and use. A *virtue* of the isomorphism approach is that it can cash content out in terms which do not appeal to use (i.e. the existence of a mapping between representation and represented). Then the functional role of a representation is left free to determine whether the representation is used effectively or not. When not, then the representation is misused (Cummins, 1996, Ch. 4).

I agree with Cummins on this point. I do not take the possibility of misrepresentation as constitutive of representation. Rather, the necessary feature for a structure to constitute a representation is that it present again information about its target. Thus, even though they cannot misrepresent, N-representations are indeed representations. Furthermore, once an N-representation is used, it may be misused, and therefore functioning N-representations in the sense discussed above may indeed “misrepresent” in the sense of being used to generate actions in a manner which is not efficacious for the organism’s ends.

Consider a thermostat with faulty wiring such that it blows warm air if the ambient temperature is warmer than its preset target temperature and cold air if ambient temperature is cooler. The bimetallic strip in this thermostat N-represents the environment just as effectively as in a properly functioning thermostat. Furthermore, the thermostat responds to that representation systematically, i.e. it is indeed a functioning N-representation. Nevertheless, once we introduce an analysis of the goal of the thermostat, we can see that it is misusing this N-representation. This diagnosis depends on our having different analyses of the representational content of changes in the bimetallic strip and of the functional role these changes play in determining the operation of the thermostat. Whether we call the position of the bimetallic strip a representation or not, it contains information about ambient temperature which may be exploited, i.e. its possible positions N-represent possible ambient temperatures.

Nevertheless, there is more to be said about the relationship between N-representations and error. Although N-representations may not misrepresent, they may represent more or less information about the target. The directional, perspectival character of N-representations discussed in Section 3.3 ensures that representations of the same type, i.e. induced by the same causal process, may contain very different quantities of information about the target. In some cases, this feature alone is enough to explain an instance of *apparent* misrepresentation.

Consider the example in Figure 1.⁶ If we accept the definition of objective similarity given above, then the same clause which ensures that footprints made by the same foot are similar implies that these two shadows are similar. The radical difference in perspective set by the two sources of illumination, however, has produced two quite different shadows of the cylinder: a square and a ring. These two shadows are not even topologically equivalent—the objective similarity which exists between them is very coarse (i.e. the relevant homomorphism preserves very little structure).

⁶This example was inspired by the cover of Douglas Hofstadter’s *Gödel, Escher, Bach*, which features wooden blocks which cast the shadow of a “G” when illuminated along one axis, an “E” when illuminated on another, and a “B” when illuminated along the third.

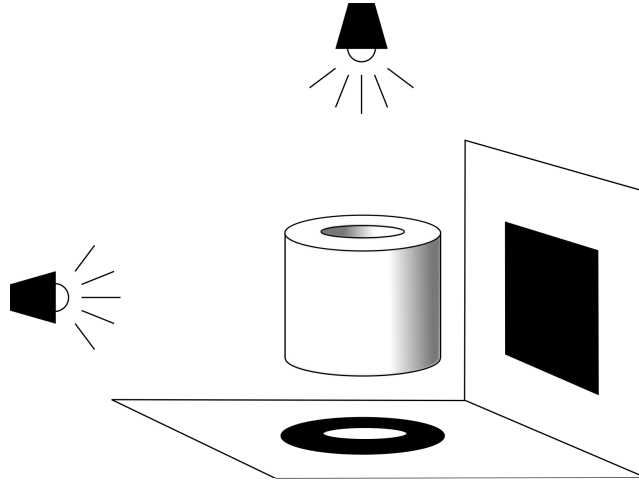


Figure 1: Perspective can produce misleading instances of objective similarity.

The shape of the cylinder *constrains* the possible shadows it can cast. In this regard, the square is very special as a cylinder shadow. The cylinder will cast a square shadow only if a light source is oriented along the horizontal plane which bisects the cylinder widthwise. As soon as the angle or position of the light source moves from that plane, the curves of the cylinder begin to enter into its shadow. A very small subset of possible perspectives will generate square shadows from a cylinder.

The crucial point here is that the perspectival nature of N-representation ensures *variance* in the possible N-representations of an object induced by a single causal process. Along with variance in the nature of these N-representations comes also variance in which features are informative and which are not. I may take a smooth footprint as indicative that the boot which made it has no tread, but it also may be the case that the low viscosity of the mud erased the treadmarks as soon as the boot was lifted. Is it appropriate to say the smooth footprint mis-N-represents the boot? No, but it does misleadingly represent the boot in the sense that surface properties of the base of a footprint can be information-bearing, but in this case they are not.

The natural variance amongst N-representations suggests the possibility that the appearance of error may emerge in purely mechanical responses to veridical information. Consider, for example, a machine with a simple perceptual system, which responds only to light and dark. It inhabits a world of cylinders and cubes, and mechanically responds to a shadow cast on its surface by generating an internal representation of a cube or a cylinder. Square shadows may be cast by either a cube or a cylinder, but cubes cast square shadows when illuminated from more perspectives than cylinders. Thus, if our machine generates cube-representations internally whenever a square shadow is cast upon it, it will generate a cube whenever its perceptual system is in the position of the right wall in Figure 1. If it is structured, this internal cube-representation will constitute an N-representation of the cylinder. From an outside perspective, this appears to be an instance of misrepresentation, even though no norms regulate the mechanical system. Strictly speaking, all that has occurred is that some features of the internal representation which are often information bearing (e.g. its topological properties) in this case fail to bear information about the target.

Many optical illusions may be explained in just this way. The Ames room, for example, manipulates

the visual system into perceiving two men who are the same height as very different in height. It does this by placing one man much further away than the other, but in a specially constructed environment that appears when viewed from a specific vantage point like a regular room, with the men equal distances from the viewer. The visual system is “fooled” by the unusual structure of the special room into assigning different heights to the two men. But it is not as if any normativity is required to generate this “error”—it is a simple consequence of the hardwiring of the visual system. Features of the internal representation which in most cases are information bearing, e.g. distance between the top of a man’s head and the ceiling, in this unusual case fail to bear information. What appears from the outside as a misrepresentation (of height) can be explained internally in terms of mere mechanical N-representation.

5 Conclusion

The goal of this paper has been to reconcile psychological practice with philosophical theory. In particular, the psychological view that mental representations can be used to reason about objects in the world because they are similar to those objects contradicts the philosophical view that similarity is irrelevant for representation. By assuming causal structure in the world, we solved two problems in sequence. First, we avoided the problem of too many similarities, defining objective similarity in terms of causally induced homomorphisms. Second, by restricting attention to only those objective similarity relations directed along the flow of causal influence, we defined a coherent notion of representation in nature, or N-representation.

In surveying psychological theories of perception, navigation, and spatial reasoning, we saw mental representations identified with functioning N-representations. On this account, the objective similarity between a mental representation and its target explains its *utility* for guiding action; the fact that the representation was caused by its target explains its *relevance* for guiding action; and the *use* of the representation for guiding action locates a source for potential error in the potential for misuse. Furthermore, we demonstrated that these considerations are orthogonal to the debate between symbolic and imagistic accounts of representation.

I hope the account of objective similarity developed here has made similarity safe again for the study of mental representation, and perhaps even for the study of representation in general. At the very least, our discussion has taken assumptions about representation implicit in the psychological literature and made them explicit, clearing the way for further debate.

References

- Aristotle (1935). *On the Soul, Parva Naturalia, On Breath*. Harvard UP, Cambridge, MA.
- Bartels, A. (2006). Defending the structural concept of representation. *Theoria*, 55:7–19.
- Churchland, P. M. (2002). Outer space and inner space: The new epistemology. *Proceedings and Addresses of the American Philosophical Society*, 76:25–48.
- Churchland, P. M. (2007). On the reality (and diversity) of objective colors: How color-qualia space is a map of reflectance-profile space. *Philosophy of Science*, 74:119–149.

- Cummins, R. (1996). *Representations, Targets, and Attitudes*. MIT Press, Cambridge, MA.
- Cummins, R., Blackmon, J., Byrd, D., Lee, A., and Roth, M. (2006). Unexploited content. In McDonald, G. and Papineau, D., editors, *Teleosemantics*, pages 195–207. Oxford UP.
- Cummins, R. and Poirer, P. (2004). Representation and indication. In Staines, P. and Slezak, P., editors, *Representation in Mind*, pages 21–40. Elsevier.
- Deutsch, J. A. (1960). *The Structural Basis of Behavior*. University of Chicago Press, Chicago, IL.
- Dretske, F. I. (1981). *Knowledge and the Flow of Information*. MIT Press, Cambridge, MA.
- Dretske, F. I. (1988). *Explaining Behavior: Reasons in a World of Causes*. MIT Press, Cambridge, MA.
- Dretske, F. I. (1995). *Naturalizing the Mind*. MIT Press, Cambridge, MA.
- Gallistel, C. R. (1980). *The Organization of Action: A New Synthesis*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Gallistel, C. R. (1993). *The Organization of Learning*. MIT Press, Cambridge, MA.
- Gombrich, E. (1960). *Art and Illusion: A Study in the Psychology of Pictorial Representation*. Pantheon Books, New York, NY.
- Goodman, N. (1972 [1970]). Seven strictures on similarity. In *Problems and Projects*, pages 437–446. The Bobbs-Merrill Company, Inc., New York, NY.
- Goodman, N. (1976). *Languages of Art*. Hackett, Indianapolis, IA, 2 edition.
- Hofstadter, D. (1979). *Gödel, Escher, Bach: And Eternal Golden Braid*. Basic Books, New York, NY.
- Kosslyn, S. M., Thompson, W. L., and Ganis, G. (2006). *The Case for Mental Imagery*. Oxford UP, New York, NY.
- McLendon, H. J. (1955). Uses of similarity of structure in contemporary philosophy. *Mind*, 64:79–95.
- Newman, M. H. A. (1928). Mr. Russell’s “causal theory of perception”. *Mind*, 37:137–148.
- Palmer, S. E. (1999). *Vision Science: Photons to Phenomenology*. MIT Press, Cambridge, MA.
- Perlman, M. (1997). The trouble with two-factor conceptual role theories. *Minds and Machines*, 7:495–513.
- Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, 25:157–238.
- Shepard, R. (1975). Form, formation, and transformation of internal representations. In Solso, R. L., editor, *Information Processing and Cognition*, pages 87–122. Wiley, New York, NY.
- Shepard, R. (1981). Psychophysical complementarity. In Kubovy, M. and Pomeratnz, J. R., editors, *Perceptual Organization*, pages 279–341. Lawrence Erlbaum Associates, Hillsdal, NJ.
- van Fraassen, B. (2008). *Scientific Representation*. Oxford UP, New York, NY.

Watanabe, S. (1969). *Knowing and Guessing: A Quantitative Study of Inference and Information*. Wiley, New York, NY.