Representation in Biological Systems: Teleofunction, Etiology, and Structural Preservation

Michael Nair-Collins

Abstract In this chapter I propose a novel thesis about the nature of representation in biological systems. I argue that what makes something a representation is distinct from what determines representational content. As such, it is useful to conceptualize *what it is to be* a representation in terms of fundamental concepts from biology, particularly the concept of a biological function (or teleofunction). By contrast, representational *content* is best understood as a structured relation involving two parts, and the explanation of how states of biological systems have content involves the preservation of internal structural relations and causal history.

I review recent literature on the neurophysiologic mechanisms underlying a sensory discrimination task, in which neurons use a variety of mechanisms for encoding, storing, and comparing information about vibrotactile stimuli. These mechanisms include a one-to-one burst code, a temporal code in which periodicity is the operative mechanism, and a variety of rate codes, some with opposite slopes, and some reflecting neither the base nor comparison stimuli, but rather their quantitative difference. In motor cortex, a binary behavioral outcome is reflected in a sigmoidal shape of firing patterns. A theory of biological representation, if it is to be empirically useful, ought to be able to unify these various encoding mechanisms under an overarching conceptual framework that explains what biological representation is and how representational content is determined, from a general standpoint, and I suggest that the theory on offer takes significant steps toward this aim.

M. Nair-Collins (✉)
Medical Humanities and Social Sciences, Florida State University
College of Medicine, Tallahassee, FL, USA
e-mail: michael.nair-collins@med.fsu.edu

L. Swan (ed.), *Origins of Mind*, Biosemiotics 8, DOI 10.1007/978-94-007-5419-5_8,

1 Introduction

Representation is a foundational concept. At its core, it is simply *aboutness*, pointing-to, or standing-in-for. For example, my belief that I have a cup of coffee on my desk is *about* the cup of coffee. As I consider what would happen were I to turn it upside down, the processes involved in counterfactual reasoning and visual imagery involve states that "stand in for" or represent the cup in different positions, the likely outcomes such as coffee spilling on my desk, and so on. This basic concept of "aboutness" is appealed to routinely – in various incarnations – in the cognitive sciences, the neurosciences, our commonsense psychology, and in the philosophy of mind and language. It is used to explain many aspects of neurological and cognitive functioning as well as adaptive (and maladaptive) behavior. Indeed, we might reasonably consider the concept of representation to be the *single* foundation upon which our understanding of mind rests. Yet it is widely agreed that we lack an adequate naturalistic understanding of representation and its place in the physical world. There is something deeply mysterious about how physical systems have states that bear this sense of "directedness," particularly given that such systems can make errors and can represent counterfactual scenarios, both of which seem to imply a relation between the representation and a nonexistent state of affairs. My purpose in this chapter is to propose a thesis about the nature of representation in biological systems.

As mentioned above, the concept of representation gets imported into a number of distinct theoretical approaches to understanding the mind/brain and behavior, from neurophysiology, to cognitive psychology, to our commonsense belief-desire psychology. My purpose in this chapter however is only to address the primitive or basic representational states instantiated in the nervous systems of living biological organisms, from which more complicated states presumably arise.

2 Representation: Accuracy, Error, and Logical Structure

Not everything in the universe is a representation. This is obvious, surely, but the question then arises as to what differentiates things that do, from things that do not, bear representational content. One of the most prominent responses given among philosophers of mind is that representations are states that are truth or satisfaction-evaluable, meaning that they are states that can be evaluated as to whether they are accurate or inaccurate, satisfied or not. For example, my belief that there is a cup of coffee on my desk is truth-evaluable; it might be accurate, or the belief might be inaccurate. Suppose I have an intention to pick up the cup; that intention might be satisfied (i.e., I might actually pick up the cup) or not. Indeed, the problem of incorporating an understanding of *misrepresentation* into an account of representation is perhaps the single most discussed problem in the last 30 years of work on mental representation in analytic philosophy.

This is a key conceptual point that bears emphasis. It is common in neuroscience to assume something like an implicit causal theory, wherein neurons or ensembles of neurons that are differentially responsive to certain forms of energy at the periphery (e.g., edge detectors in primary visual cortex) are taken to represent what typically causes them to fire (e.g., bars of light at a particular orientation relative to the retina; cf. Bechtel 2001, for discussion). Farther downstream, other neurons are assumed to take the information encapsulated in the firing of edge detectors with affinities for specific orientations and generate progressively more complex and abstract representations of visually encoded objects.[1] However, simply differentially responding to (i.e., being caused by) specific kinds and levels of energy is not sufficient for something's having representational content.[2] A tropical storm system, for example, is differentially responsive to specific causal factors involving atmospheric pressure and temperature, wind speed and direction, and so forth. But the states of that system do not bear representational content, and there is no sense in assigning to them semantic properties such as accuracy or error. If a state of a system is not truth- or satisfaction-evaluable, then there is no distinction between its simply having a causal history or playing some causal role (which everything does) and its being a representation, being an encoding, bearing representational content, etc. Representations, of course, play causal roles as well, but they are also semantically evaluable; indeed, this is what generates the mystery in the first instance.

It also bears emphasis that the concept of truth-evaluability is not specific to human languages or linguistically expressed beliefs and desires. Presumably, the honeybee's dance represents the location of nectar to its conspecifics, with variables on the dance structure such as tempo and the angle of its long axis corresponding to variables on nectar location such as distance from the hive and direction relative to the sun (cf. Millikan 1984, chapters 2 and 6). Such dances are semantically evaluable: The dancing bee can send its conspecifics directly to the nectar by accurately representing its location, or it can send them in the wrong direction by misrepresenting the location of nectar.

Similar comments can be made regarding early perceptual and discriminatory processes: Rats are able to use their whiskers to discriminate the size of an aperture in order to select one of two options that will lead to their acquisition of a food pellet in a laboratory task (Nicolelis and Ribeiro 2006; cf. Swan and Goldberg 2010). Supposing the animal's trained task is to press the left button when the aperture is narrow (vs. wide), the animal might be in error by pressing the right button instead. In this case, its behavioral signal is incorrect; this might be a result of any number of

[1] The classic "hierarchical processing" view of visual representation adumbrated above is of course complicated by the fact that feedback modulation occurs at every hierarchical level, even prior to primary visual cortex (V1) in the lateral geniculate nucleus of the thalamus. But that does not alter the basic conceptualization of the representational capacity of early sensory neurons as being grounded in a specific causal etiology.

[2] This is not to say that edge detectors are, or are not, representational; rather, it is to say that if they are, it is not solely in virtue of their affinity for firing in response to certain types of energy impinging on the periphery.

factors. Its early perceptual encoding and discriminatory processes might be in error by encoding the width as wide when it is in fact narrow; its short-term memory might be in error by losing the informational content from early perceptual processes as it transforms sensory and mnemonic information into motor plans; its long-term memory might be inaccurate by reversing the task instructions (e.g., recalling the task instructions as to press the right button for narrow rather than the left); and its motor command processing might generate a motor output different than what the system had intended (e.g., pressing the right rather than left button). Each of these would lead to the behavioral manifestation of task error in a given trial. But each of these states, from perceptual discrimination to short- and long-term memory, to motor plans, to behavioral output, is semantically evaluable in the sense that it can be accurate or inaccurate (for sensory and mnemonic representations as well as behavioral signals of the choice made), or satisfied or not (for motor plans). By contrast, the states of a tropical weather system, though such systems are nearly as causally complex, are not amenable to such interpretation and are not representational. *Representational content demands the possibility of accuracy or error*, and this has a significant consequence for a theory of representation in biological systems.

For a state to bear representational content and hence be truth- or satisfaction-evaluable, it must be logically structured. A linguistic example is instructive here. The sentence, "Johnny has green hair," let us assume, is true. It is true in virtue of (1) the subject term "Johnny" refers to, or points to, Johnny, thereby rendering the sentence itself as referring to Johnny, and (2) the predicate term "has green hair" predicates the property of *having green hair* of whatever thing the sentence refers to. In this case, that thing is Johnny. Furthermore (we are assuming), Johnny does indeed instantiate the property of having green hair, and therefore, the sentence is true; if he did not, then the sentence would be false. This basic linguistic distinction between subjects and predicates maps onto the ontologically basic distinction between individuals and the properties that they bear, with subjects referring to individuals and predicates applying to properties. I'll henceforth refer to the relation between subject and object as *reference* and the relation between predicate term and property as *predication*. It is crucial to recognize that subjects, or referential terms, in and of themselves, are not truth-evaluable, and neither are individual predicates truth-evaluable. The term "Johnny" is neither true nor false, and the term "has green hair" is neither true nor false. It is only their concatenation, or joining together in a unified semantic construct, that renders truth- or satisfaction-evaluability, and hence accuracy or error, possible. Thus, neither reference nor predication alone, in the absence of their logical concatenation, suffices to generate representational content. Representational content demands the possibility of accuracy or error, as discussed above, and accuracy and error do not occur except in the context of a representation that bears logical structure.

This concatenation, or logical structure, need not imply physico-mechanical or symbolic structure. In natural languages, the logical structure of sentences supervenes on its syntactic structure, itself realized by orthographic or phonetic structural properties (for written and spoken utterances, respectively). However, even apparently (physically) unstructured entities can bear logical structure. By "logical structure" I mean simply that the vehicle of representation both refers to a thing

and predicates a property of that thing. An example that Devitt and Sterelny use in discussing whether representations can be simple is the yellow flag once hung on a ship's mast to signify to other passing ships that the ship has yellow fever (Devitt and Sterelny 1999, 139). This seems like a simple, nonstructured vehicle of representation, but it isn't, at least not in the sense that I'm using the term. The fact that the flag *is yellow* signifies that whatever ship is flying it has yellow fever. But it is not the yellowness of the flag that signifies *which ship* has yellow fever. The fact that the flag is attached to *this* ship's flagpole is what determines the referent of the predicate, "has yellow fever," as this particular ship. Thus, different aspects of a vehicle of representation can determine different aspects of its representational content; logical structure can, but need not, map onto physical or symbolic structure.[3]

Building on this background, I'll next briefly outline a proposed theory of representation in biological systems, followed by an illustrative example appealing to recent work on the neurophysiological mechanisms involved in Macaque (and by extension human) vibrotactile discrimination.

3 A Theory of Representation: Teleofunction, Etiology, and Structural Preservation

There is a conceptual distinction between what makes something a representation and, given that a thing is a representation, what determines its content. The former involves the metaphysics of what it is to be a representation, and the latter, the

[3] The argument I'm building here is that the fundamental semantically evaluable units are themselves truth-evaluable; hence, those units bear logical structure in the sense I'm using the term here. A different possibility is that the basic semantically evaluable units are not themselves truth-evaluable, but are instead something like subsentential units that concatenate to form larger sentence-like, truth-evaluable complexes. These fundamental units are like words in a language of thought, admitting of syntactic rearrangement which generates the productivity and systematicity of the language of thought, itself responsible for the productivity and systematicity of natural languages (Fodor 1975, 2008). This is the (or at least one of the) standard view(s) in classical cognitive science. However, the key step is the concatenation of numerically distinct, neurologically instantiated symbols: How does it work? How and why do those two neurologically instantiated symbols "come together" in that particular thought, and not some others? In virtue of what is this complex well-formed in its neurological syntax? In virtue of what are these symbols "joined together"? The appeal to concatenating neurologically instantiated symbols at the lowest level introduces a new binding problem: How and why do those particular symbols join together, excluding others, and in what does this joining consist? Just like the more familiar binding problem of explaining how different aspects of an experience (e.g., bluishness and squareness) join together in the brain to form a coherent, unified percept (e.g., as of a blue square), the *syntactic binding problem* demands an explanation for how distinct symbols join together to form a unified meaningful mental representation. If, however, the fundamental semantic units are, as I suggest, themselves logically structured and hence truth-evaluable, then the syntactic binding problem is avoided for those units. Furthermore, many suppose that even the lowest-level sensory states can *accurately* or *inaccurately* reflect peripheral energy states. If that is the case, it follows that the sensory states must have logical structure because neither accuracy nor inaccuracy is possible without it, as argued in the text. There is of course a great deal more to be said on this issue, but I will leave further discussion for a different venue.

semantics of representational content. To compare, consider the difference between what makes something money and, given that a thing is money, what determines its particular value (Michael Levin proposed this analogy in conversation). Although the conditions that determine each are closely related (involving complex relations and interactions among social agents), there is nonetheless a conceptual distinction between a thing's being money and, given that it is money, what its particular value is. For example, the value of a dollar, understood in terms of its relative purchasing power either locally or globally in exchange for foreign units of currency, fluctuates. But its status as *being money* (at all) does not; therefore, they are conceptually distinct.

This distinction is helpful in the present context as follows. Representations are states of biological organisms. As such, it is useful to conceptualize *what it is to be* a representation in terms of fundamental concepts from biology, particularly the concept of a biological function (or teleofunction). Just as hearts have the function of circulating oxygenated blood, but can fail to do so, representational states of the nervous system also have biological functions to play (but can fail to do so). Living, mobile organisms have the capacity to selectively respond to labile environmental conditions – in ways that reflect those changing conditions – which enables them to maintain physiologic stability, to avoid predation, or to reproduce. The behavioral flexibility that manifests as appropriate responses to changing environmental conditions is rooted in the organism's capacity to represent both internal and external conditions; more specifically, *what it is to be* a representation is to have the biological function of bearing certain correspondence relations, as follows.

Some things have the biological function of corresponding to environmental conditions in such a way that other states, the *users* or *consumers* of the first, use the state of the first in reacting appropriately to changing internal or external conditions. Other things have the biological function of producing or helping to produce the states to which they correspond. The former are indicative or sensory representations, and the latter are procedural representations or motor plans (cf. Millikan 1984, 1989, 2004).

For example, the nematode *C. elegans* performs chemotaxis, or oriented movement in response to a chemical stimulus, to locate its primary food source of bacteria. The chemotaxis circuit includes four pairs of chemosensory neurons, four pairs of interneurons, and five pairs of motor neurons (Bargmann and Horvitz 1991). *C. elegans* neurons exhibit graded voltage potentials (rather than action potentials); the voltage of the chemosensory neurons at the tip of its nose bears specific correspondence relations to the concentration of chemoattractant in the environment, whereby increases in chemical concentration correspond to proportional increases in voltage. By comparing the scalar value of the current chemical environment to its first derivative (i.e., the change in chemical concentration as the nematode moves), the sensory and interneurons generate a signal to the motor neurons, which then generate a motor output signal to the neck muscles, enabling the animal to orient itself up the chemical gradient and toward food (Ferree and Lockery 1999; cf. Mandik et al. 2007, for computer simulations of evolved neural network control of chemotaxis). In this example, the sensory neurons have the

teleofunction of bearing specific correspondence relations to the concentration of chemoattractant at the periphery of the organism. In virtue of the sensory neurons realizing this correspondence relation, the interneurons and motor neurons are able to use that information to generate output signals appropriate to the local environment by comparing the present concentration to the change in concentration in order to determine in which direction the gradient increases. Thus, the changing voltages of the chemosensory neurons are sensory or indicative representations. The motor neurons evince similar proportional changes in voltage relative to the degree of extension of specific muscles in the neck which determine the neck's turning angle, and the activity of the motor neurons is causally relevant to producing those specific turning angles. Thus, these neurons have the function of producing (or causing) the muscle states to which they correspond, and should be considered procedural representations or motor plans. What it is to be a representation, therefore, is to have the biological function of bearing specific correspondence relations which enable adaptive behavior of the organism of which those states are a part.

However, as discussed in the previous section, representational *content* demands the possibility of accuracy or error, which in turn requires logical structure. Having the biological function to bear specific correspondence relations to environmental or muscle states is insufficient for generating logical structure, and thus is insufficient for generating representational content. In order to explain what determines representational content (as opposed to what makes a thing a representation at all), some analogue of predication and reference must be built in to the theory. I emphasize again that these concepts are not specific to language, but instead map onto the basic ontological distinction between properties and the bearers of properties. Even the representational states of worms – if they are to bear representational content and thus admit of accuracy and error – must both refer to a thing and predicate some property of that thing. The states of the chemosensory neurons of *C. elegans*, for example, might predicate *having concentration X of chemoattractant* (a property) of the immediate environment located at the tip of its nose (a *thing* of which the property is predicated). Of course, the worm does not use words like "concentration," "chemoattractant," or "local environment" to *express* such representational contents, but this does not imply that its neural states do not thereby *have* that representational content.

I propose that what determines representational content is a combination of causal etiology and isomorphism. As discussed above, it is common in neuroscience to implicitly presume some version of a causal theory of representation, whereby states of the nervous system are taken to represent what typically causes them, or what they typically cause. Although this is an insufficient condition on being a representation, it is nonetheless a key component of a theory of representational content. However, it is also well understood from the philosophy of mind literature how profoundly difficult it is to make sense of the possibility of error, given a purely causal theory of representational content.

There are two kinds of causal theories: causal history (or etiology) and counterfactual covariation. Causal history theories state that representations represent

whatever caused them. In this circumstance, it should be obvious that error is impossible: Representation R represents precisely its causal antecedent; therefore, no sense can be had in stating that the representation is in error. The frog that snaps after a passing bit of darkly colored leaf blowing erratically in the wind, which resembles a fly, cannot be said to have misrepresented the leaf as a fly. Instead, it must be said that the frog correctly represented the leaf, but then it is difficult to make sense of why the frog snapped at it. To deal with such problems, the concept of counterfactual covariation was introduced, in which representational states are taken to represent whatever they counterfactually causally covary with, perhaps under ideal circumstances, or ideal circumstances in the environment of evolutionary origin. But a different set of problems then arise, the most significant of which is that attempting to discern the item or property of maximal counterfactual covariance inevitably leads to a disjunction of such things and thereby, again, the impossibility of error. For example, the states of the frog's nervous system which are typically taken to represent the fly as food do not maximally covary with flies, but rather with the disjunctive property *fly-or-passing-leaf*. In this circumstance, error is again impossible because the frog correctly represents the passing leaf as *fly-or-passing-leaf*, but it seems clear that we should say that the frog has mistaken the leaf for a fly. That's why the frog snapped at it.

However, there is also wisdom in causal theories, which (I suspect) is why they are implicitly presumed in the neuroscience literature and why so much energy has been expended in the philosophical literature to attempt to correct their serious deficiencies. To appreciate why causal etiology is relevant, consider the parallels between reference and causation. The basic problem with causal theories is that a causal relation either obtains or does not, and if it does, it becomes very puzzling to say why in some circumstances, but not others, this causal relation should determine representational content. But reference (alone), like causation, either obtains or does not. There is no such thing as "mis-reference"[4]; semantically evaluable units must either succeed or fail in referring (to anything). Thus, while we cannot reduce representational content to causal etiology because of the impossibility of error, we can reduce *reference* to causal etiology, without needing the possibility of "error." Referring expressions are neither true nor false; rather, they either refer or they don't. In explaining reference in terms of causal etiology, however, it should be understood that causal history determines the object or thing that the representation

[4] We'll need to be careful here: If I "refer" to my dog Mac as "that cat," it might seem that I've mis-referred, but I haven't. Rather, the ostensive act referred to an individual, and I predicated the property *catness* of it. The reference relation obtained, whereas I misapplied a predicate of that to which I referred. On the other hand, there are tricky issues regarding reference to nonexistents; can I refer to Sherlock Holmes or unicorns? These are larger issues in the philosophy of language which will not be addressed here; better to understand the simpler kinds of representation first. If you like, consider my claim that there is no mis-reference as both axiomatic and using the word "reference" to mean something like, only the most fundamental kind of reference. The argument for accepting any axiom is, of course, dependent on how well the theory constructed from that axiom works.

is about, but does not determine the property that the representation predicates of that thing.

A different and much older idea says that representation is a picturing or resemblance relation, where the vehicle of representation bears structural similarities to, or shares properties with, that which it represents. The guiding idea here is that there is a kind of resemblance or "mirroring" between representation and represented in virtue of which the representation relation obtains. The strength of this view is its intuitive appeal: A realistic portrait of President Obama represents President Obama himself, due to the structural similarities, or the resemblance, between the two. However, due to a number of problems with a simple resemblance view, among them that resemblance is symmetric while representation is not, resemblance was abandoned long ago as a viable theory of representation. It has lately been revived, however, by appealing to a more sophisticated form of resemblance, namely, an isomorphism among a *system* of representations and a *system* of states of affairs, rather than a structural similarity between the token vehicle of representation and whatever it represents.

On this latter theory, the guiding motivation is the same: The preservation of internal structural relations between representation and represented is of the essence of representation. However, the structural similarity obtains between a set of items and relations on that set, and another set of items and relations on it. By appealing to systems of states of representational vehicles and transformations over them, a more abstract kind of resemblance can obtain, which need not respect any first-order structural similarity between a token vehicle of representation and its content. This is important because for the most part, a first-order picturing or mirroring relation does not hold between brain states and world states (e.g., the chemosensory neurons of *C. elegans* do not share first-order structural similarities with the changing chemical concentration at the tip of its nose, in the same way that a realistic portrait of President Obama shares a first-order structural similarity with President Obama himself).

While the system isomorphism approach is in many ways an improvement over its ancestor, it still faces many of the same problems. The most important of these is the problem of multiple isomorphisms. If isomorphism is the sole determinant of content, then it seems to follow that representations are about or represent far too many things. For example, given any relational system (i.e., a set with relations on it), there exist infinitely many relational systems to which it is isomorphic; furthermore, given two isomorphic systems, there exist numerous if not infinitely many distinct mappings between those two systems that preserve isomorphism equally well. Apparently, this would seem to preclude the possibility of false representations since a representation may be true under one mapping but false under another, and if there is no principled means of selecting among the numerous mappings, then there seems no way to account for error.

Consider, however, the parallels between predication and isomorphism. Unlike causation, and unlike reference, predication is not specific. The predicate "has green hair" applies to all and only the things that have green hair; predicates are multiply applicable because properties are multiply instantiated, unlike individuals which are

not. Unless concatenated with a referential expression, a predicate does not apply to any specific individual. But notice that this is precisely the problem with isomorphism-based theories: They are not specific. The multiplicity of isomorphisms, and the multiplicity of things to which predicates apply (due to the multiple instantiability of properties), suggests that isomorphism or something like it is the element responsible for predication in basic representations.

More specifically, states of individual neurons or ensembles of neurons admit of certain transformations that realize an ordering relation over those states, resulting in empirical relational systems. Firing rate, for example, admits of transformations by increasing or decreasing how quickly action potentials fire; the set of firing rates ordered by the greater-firing-rate relation constitutes an empirical relational system. Similar remarks apply to neurons that admit of graded voltage potentials, ordered by the greater-voltage relation. Furthermore, transducible energy states impinging on the periphery of an organism can be ordered according to transformations in similar fashion, resulting in relational systems composed of distinct energy states and transformations over them. For example, the set of concentrations of chemoattractant in the local environment can be ordered by the greater-chemoattractant relation, resulting in a relational system. The idea is that representations are not found in biological organisms as punctate atoms, but rather there are *systems* of representations, the members of which are organized in such a way that those systems are isomorphic to different organized systems of representeds. A mapping, or mathematical function, from the elements of one system to the elements of the other maps states of one system (say, a particular firing rate) to states of the other (say, a particular frequency of vibration at the skin) so that that particular firing rate predicates the property of vibrating at that particular rate. This mapping just is the specific correspondence relation mentioned above, which these representational states have the teleofunction of bearing.

Furthermore, there is no need to constrain this idea to the activity of single neurons. Populations of neurons can be described using vectors and relations on them, and multivalued functions between higher-order relational systems and other relational systems describing energy states can define isomorphisms between systems. On the represented side, anything can be a member of a relational system, not just parametric energy states at the periphery of the organism. Thus, in addition to mechanical, electromagnetic, thermal, and other forms of energy, relational systems may include things like predator, food source, conspecific, shelter, etc. There is also no reason to suppose that the mappings between relational systems must involve linear or even monotonic relations.[5] They can be sigmoidal, quadratic, or anything at all. Finally, for marshaling the concept for use in a theory of biological representation, there seems no reason to maintain the relatively strict technical requirements imposed by

[5] Akins (1996), for example, argues that the "traditional naturalist" project of Dretske (1981, 1988), Fodor (1987, 1990), Millikan (1984, 2004), and others rests on a mistaken view of the senses, which is that they must be "veridical." Akins argues instead that sensory systems are not veridical but are what she calls "narcissistic." That is, they do not "dispassionately" report what is going on out in the world, but instead are highly dependent on local context (as in, "what does this mean for

the mathematical construct of isomorphism. There are numerous ways to extend or relax these technical constraints while maintaining the fundamental aspect of the preservation of internal relational structure (for some examples, see Swoyer 1991). I use the term *structural preservation* to refer to the class of structure-preserving relations between relational systems that includes isomorphism, homomorphism, and several others, which are weakened versions of these constructs.

Before delving into a detailed example to illustrate the theory, I'll summarize the main ideas. Not everything is a representation; what differentiates things that are, from things that are not representations, is semantic evaluability, which requires the possibility of accuracy or error. This applies to even the simplest biological organisms, not just language-using humans. Furthermore, the possibility of accuracy or error requires logical structure, or a concatenation of some analogue of reference and predication, where reference maps subjects to objects (or things) and predication maps predicates onto properties. However, logical structure need not imply physico-mechanical or symbolic structure; rather, different aspects of a vehicle of representation might be responsible for the different aspects of representation.

There is a conceptual distinction between what makes something a representation and what determines representational content; a theory of representation must explain both. I've suggested that what makes a thing a representation (at all) is its having the teleofunction of bearing certain correspondence relations which enable the organism to respond appropriately to changing environmental conditions. However, to explain representational content, an explanation of both reference and predication is required (because logical structure is required), and the teleofunctionally determined correspondence relations are, by themselves, insufficient to explain both components. However, I've suggested that causal etiology is the aspect of a representational vehicle that determines the thing to which it refers. Furthermore, isomorphism between systems of representations and systems of represenetds determines the specific property predicated of the thing to which the representation refers. The correspondence relations that the state has the teleofunction of bearing to energy states at the periphery just are the mappings between relational systems that determine isomorphism and match up, one-to-one, states of the representational

---

me, the receptor?"). This objection is somewhat strange in that what *constitutes* veridical representation is precisely the question. Thus, in order to say that sensory systems are not veridical, one must first be committed to some theory of representational content. Her claims that thermoreceptive systems are not veridical, therefore, cannot be used as an objection to the very project of understanding veridicality itself. Akins, apparently, considers thermoreceptors and the neural machinery attached to them to be narcissistic and non-veridical because they do not have linear response profiles, but instead have very complicated response profiles depending on local context. This doesn't show that they are not veridical, just that they behave according to complicated nonlinear correlations to the environment, and can change in different contexts. These complicated response profiles nonetheless describe mapping functions between relational systems composed of neural activity and relational systems composed of energy states, and bearing these response profiles may very well be what these thermoreceptors and other neural machinery are *supposed to do*; that is, have the teleofunction of doing.

system (e.g., specific voltages) to states of the represented system (e.g., specific concentrations of chemoattractant). Henceforth, I'll refer to the theory as the *structural preservation theory* of representation.

4 The Neurophysiological Mechanisms of Vibrotactile Discrimination

In what follows, I describe a research program aimed at delineating the neural and cognitive mechanisms that underlie vibrotactile discrimination. I then use these results to illustrate the structural preservation theory of representation and furthermore to show how the theory helps in interpretation of the empirical results. The basic, classical task (LaMotte and Mountcastle 1975; Mountcastle et al. 1990) is as follows. A seated Macaque monkey has its left hand secured, palm up. A stimulator tip is lowered, indenting the skin of one of the monkey's fingertips; it is not vibrating at this point. The monkey then presses a key with its free right hand and holds the key down. The stimulator then produces a sinusoidal vibration, between 5 and 50 Hz, to the left hand fingertip (this is the *base stimulus*, or f_1 for first frequency), followed by a delay period (or *interstimulus interval*), followed again by a second vibration (the *comparison* or f_2), also between 5 and 50 Hz. At the offset of the comparison stimulus, the monkey releases the key with its right hand and signals its choice on which frequency was faster by pressing one of two push buttons located at eye level. The monkey is rewarded with a drop of juice for correct discrimination.

A schematic of the neural events that occur during this task is as follows. Rapidly adapting, superficially located mechanoreceptors in the finger known as *Meissner's corpuscles* transduce the mechanical energy into action potentials, which travel up the spinal cord, through the thalamus, into primary somatosensory cortex (S1), and thence to the secondary somatosensory cortex, or S2 (Gardner and Kandel 2000; Gardner et al. 2000; Vallbo 1995). The outgoing signal from S2 then gets widely distributed, to at least the prefrontal cortex (PFC), the ventral premotor cortex (VPC), and medial premotor cortex (MPC); PFC and VPC both appear to be serially connected to MPC. Then MPC transmits activity to the primary motor cortex (M1), whose activity ultimately results in the monkey's button-pressing behavior signaling its choice (Romo et al. 2004a). These cortical areas are typically associated with cognitive activities as follows. Primary and secondary sensory areas are involved in sensory processing. PFC is widely implicated in short-term or working memory processes, and MPC/VPC are considered to be premotor areas, which begin the transformation of signals from sensory and memory processes into motor plans. Primary motor areas are associated with the implementation of generalized motor plans, which then get refined into more specific muscle commands, taking into account various feedback mechanisms by the basal ganglia, cerebellum, and spinal cord.

The neural activity that occurs during the presentation of the stimulus is as follows. In the periphery, neural firing is phase-locked to the stimulus, where the neuron fires a spike or burst of spikes for each amplitude peak of the sinusoidal stimulus (Mountcastle et al. 1969, 1990; Salinas et al. 2000). Traveling into the cortex, there appear to be two subpopulations in S1.[6] In the first, subpopulation-1, neural activity is no longer phase-locked to the stimulus, but the temporal structure of neural firing correlates with the stimulus frequency, as follows. Periodicity is the property of exhibiting regular, repeating characteristics. Using a Fourier decomposition of the firing pattern, it is possible to deconstruct the function describing that pattern into its component sine and cosine functions, as well as determine their "power," or determine which frequency contributes most to the original function. In subpopulation-1 of S1, the power spectrum frequency at peak (*PSFP*), which is the frequency that contributes most to the firing pattern, matches the frequency of the tactile stimulus (Hernandez et al. 2000; Salinas et al. 2000). In subpopulation-2 of S1, the firing pattern becomes less periodic, and the PSFP is no longer matched to the frequency of the stimulus. However, the aperiodic firing pattern now correlates with stimulus frequency in terms of its rate, approximating a monotonic linear function of rate (Salinas et al. 2000).

In S2 and beyond, the rate correlation remains prominent, and the temporal, periodicity-based, or phase-locked code is no longer evident. An important difference emerges in S2. As in S1, there are subpopulations characterized by their differential responses to sensory stimuli; however, in S2 and in all of the more central areas of this circuit, the subpopulations are oppositely "tuned" (Salinas et al. 2000; Romo et al. 2004a). In S1, all neurons increase their firing with increases in stimulus frequency. In more central areas, approximately half increase firing rate as a monotonic increasing function of increasing stimulus frequency, whereas the other half decrease their rate as a monotonic decreasing function of increasing stimulus frequency. Thus, as stimulus frequency gets slower, the negatively tuned neurons increase their firing rate. Oppositely tuned subpopulations responsive to sensory stimuli are found in S2, PFC, VPC, and MPC (Romo et al. 2004a).

The above events occur during the presentation of the base and comparison stimuli. During the interstimulus interval (of 3–6 s, although this can be increased to 10–15 s without a significant difference in performance), no stimuli are presented. To successfully discriminate the first from the second tactile stimulus, and decide which

[6] The primary somatosensory cortex is composed of four areas: 1, 2, 3a, and 3b. Each area has a complete topographic map of the body's surface composed of the receptive fields of the respective neurons. Further, the specialization of peripheral fibers seems to continue in S1; neurons are classified in S1 as rapidly adapting, slowly adapting, or Pacinian, because their firing activities are similar to their respective primary afferents (Romo and Salinas 2001, 109). The areas associated with the rapidly adapting circuit here under consideration are areas 1 and 3b. *Within* those areas, there are subpopulations, one of which appears to encode stimulus information using a temporal, periodicity-based code (described in the text), and the other using an aperiodic firing rate code (also described in the text). The terms 'subpopulation-1' and 'subpopulation-2' should not be confused with areas 1, 2, 3a, and 3b. The subpopulations here under consideration are defined by their behavior in this task and are subpopulations of anatomical areas 1 and 3b.

has a greater frequency, the animal must maintain something like a mnemonic trace of the first stimulus. During this period, neurons in PFC correlate their firing rate with the frequency of the base stimulus, with approximately half showing a monotonic increasing relationship to frequency and the other half showing a monotonic decreasing relationship (Romo et al. 1999). Correlated neural responses during the delay period are also found in S2, VPC, and MPC, also with oppositely tuned subpopulations (Hernandez et al. 2002; Romo et al. 2004b; Salinas et al. 1998, 2000).

The comparison stimulus is then presented, whereby neural activity correlates as before in terms of phase-locking and periodicity in the periphery and early S1, and transformed into a rate code in S1 and then S2. Rate is also correlated with the stimulus in PFC, VPC, and MPC. Additionally, something like a comparison and decision process now occurs, whereby the system decides which of the two frequencies is greater. The relationship of firing rate R to the base and comparison frequencies is given by the regression equation (Hernandez et al. 2002; Romo et al. 2002, 2004a):

$$R = a_1 f_1 + a_2 f_2 + c,$$

where c is a constant, f_1 and f_2 are the frequencies of the base and comparison stimulus, respectively, and a_1 and a_2 are coefficients that determine the strength of the relationship between R and frequency. When either of the coefficients is zero, there is no detected correlation between rate and that coefficient's frequency. Importantly, when $a_1 = -a_2$, then firing rate is now correlated with neither f_1 nor f_2, but with the *difference*, $f_2 - f_1$.

During the comparison period, neurons in S1 only show correlation to f_2 throughout the stimulation period; hence, the neural activities act as sensory representations of the comparison frequency. In S2, some neurons begin the period correlated with f_2, then the population as a whole shifts towards correlation with the difference, $f_2 - f_1$ (i.e., $a_1 = -a_2$) (Romo et al. 2002). In VPC and MPC, there are several different populations. Some neurons begin the comparison period correlating with the base frequency; thus, they are something like mnemonic traces, whereas others begin the period correlating with the comparison frequency as if they were sensory representations. Toward the end of the comparison period, the majority of the responsive neurons in MPC and VPC correlate with the difference, $f_2 - f_1$ (Hernandez et al. 2002; Romo et al. 2004b). Additionally, firing rates correlated with $f_2 - f_1$ are found in PFC (Romo et al. 2004a).

As with neural activity that correlates with the base or comparison frequency, the neural responses correlated with $f_2 - f_1$ (in S2, VPC, MPC, and PFC) show opposite slopes, where approximately half fire more strongly when $f_2 - f_1$ is positive, and the other half fire more strongly when $f_2 - f_1$ is negative.

Finally, M1 plays a crucial role in the animal's behavior during this task. While M1 shows no significant response above baseline activity during the base stimulus, delay period, or early in the comparison period, it does show neural activity correlated with $f_2 - f_1$, similar to the activity found in earlier areas, with subpopulations differentially responsive to the case where $f_2 > f_1$ and where $f_1 > f_2$ (Romo et al. 2004a).

In a different task, monkeys must categorize rather than discriminate the same type of tactile stimuli, simply saying whether a stimulus belongs to arbitrary categories of

high or *low* which were learned during training (Salinas and Romo 1998). In this instance, firing rates had a sigmoidal shape: For a neuron that "preferred" higher speeds, its firing rate was essentially the same for stimulus speeds of 22–30 Hz. For a neuron that "preferred" lower speeds, its rate was essentially the same for stimulus speeds of 12–20 Hz (see Salinas and Romo 1998, figures 3 and 4). Thus, as found earlier, there are two subpopulations, each of which is selective for either high or low speeds. The sigmoidal shape of the firing rate as a function of tactile speed suggests that these neurons correlate with arbitrary, learned categories ("high" or "low"). Whether or not that analysis should be applied to the tactile discrimination task is uncertain. However, M1 does appear to play a role in the decision procedure for at least the categorization task, and it does have differential activity selective for the different decisions the animal may make (i.e., base greater than comparison or vice versa). Whether that differential activity participates in the comparison and decision procedure, or simply receives a copy of a decision already made, is unclear.

5 Applying Structural Preservation Theory

It should be apparent from the above discussion that neurons in this circuit use a variety of mechanisms for encoding information about the stimuli. From the periphery and centrally inward, neurons use a simple one-to-one burst code, followed by a temporal code in which periodicity is the operative mechanism, followed by a variety of rate codes, some with opposite slopes, and some reflecting neither the base nor comparison frequency, but rather their difference. In motor cortex, a binary outcome (pressing the medial or lateral button) is reflected in the sigmoidal shape of the firing patterns. A theory of biological representation, if it is to be empirically useful, ought to be able to unify these various encoding mechanisms under an overarching conceptual framework that explains what biological representation is and how representational content is determined, from a general standpoint. I suggest that structural preservation theory does do this, mostly as a result of the versatility of the concept of isomorphism and, more broadly, structural preservation.

The first step is to establish *that* these neural mechanisms are representations; this aligns with what I've called the metaphysics of representation, or, what makes something a representation at all. I've argued that a state is a representation if it has the teleofunction of bearing certain correspondence relations such that its doing so is adaptive for the organism of which that state is a part. I'll only discuss this question with respect to burst rate in the periphery since the arguments are both simple and immediately applicable to the other neural areas and firing patterns.

The tactile sensitivity of the glabrous areas of primate skin makes possible various evolutionarily adaptive behaviors, such as grasping objects and tactile recognition, which in turn aid us in getting food into our mouths. We primates do all sorts of things with our hands, which contribute to behavior that is conducive to survival and procreation. Furthermore, the kinds and levels of energy needed to activate this circuit are very specific. Due to the microanatomy of Meissner's corpuscles, only vibrating mechanical energy in the 5–50 Hz range, at the superficially located level

(around 500 μm beneath the surface), will generate trains of action potentials. Faster or deeper vibrations simply won't activate the Meissner's circuit, but will instead activate Pacinian corpuscles, and slower indentations in the form of constant pressure will activate the slowly adapting mechanoreceptors and their associated afferents (Gardner et al. 2000; Gardner and Kandel 2000). And these are each forms of tactile, mechanical energy. Electromagnetic, chemical, thermal, or acoustic mechanical energies won't activate this circuit at all. While we should always be wary of just-so stories about evolution, it is reasonable to presume that burst rate covaries with vibrotactile frequency because, in the course of evolutionary history, there was selection for peripheral nerves that emitted a burst at a rate equal to frequency of a sine wave of pressure on the fingertip, for the specific frequency and depth ranges mentioned above, at specific anatomic locations. Therefore, the teleofunction of the primary, secondary, and tertiary afferents associated with the rapidly adapting circuit is to covary with mechanical deformations at their respective receptive fields, according to the simple function $r_1: A \rightarrow B$, where A consists of vibrotactile frequencies, B consists of burst rates, and $r_1(x)=x$. This function maps frequencies to rates, where x Hz vibrotactile frequency maps to x bursts/s. A similar argument applies to the other correspondence relations defined by periodicity and rate; therefore, they are each representational states of the organism. However, the explanation of representational *content*, allowing for accuracy and error, is given in terms of causal etiology and isomorphism.

I'll discuss four different kinds of sensory representations: the peripheral burst code, the periodic/temporal code in subpopulation-1 of S1, and both the positively and negatively sloped rate codes in S2 and beyond. We begin by defining some simple mathematical functions and relational systems. These functions are the empirically discovered correspondence relations between neural activity and ambient energy, which serve two purposes in the theory. First, these are the correspondence relations that the neural states have the teleofunction of bearing to external states; by bearing these correspondences that reflect the varying states of ambient energy, other neural processing mechanisms are able to use that correspondence to compute appropriate behavioral responses. These patterns of neural firing are representations in virtue of having the teleofunction of bearing these correspondence relations. Second, the mapping functions between relational systems define isomorphisms between those systems and match up states of neurons with energy states at the periphery, serving to determine predication. Further, as mentioned previously, for any two isomorphic relational systems, there always exists numerous if not infinitely many mapping functions between them that preserve structure equally well. However, the empirically discovered correspondences serve to rule out every other transformation on the mapping function, thus avoiding one of the key problems for isomorphism-based theories of representation.

Relational systems consist of sets with relations on them. Let $\mathfrak{A}$=the stimulus relational system and $\mathfrak{B}$=the physiological relational system, in each case that follows. Each relational system is an ordered pair consisting of a set (or domain) and a relation on that set. Hence, $\mathfrak{A}=\langle A, r\rangle$, with r being a relation on A, the domain of $\mathfrak{A}$. Isomorphism is defined by defining a bijective

function[7] from the domain of one relational system to the domain of the other, such that the relational structure of one system is preserved in the other (though the relations themselves need not be the same).[8] The domain of the stimulus relational system, *A*, consists of vibrotactile frequencies and is ordered by $>_A$, the empirical higher-frequency-than relation. The domain of the first physiological relational system, *B*, consists of burst rates. We define a *burst* in terms of interspike intervals: A burst is "a group of spikes in which all intervals between consecutive spikes [is] less than τ msec" (Salinas et al. 2000, 5504). The shorter that τ gets, the closer burst rate will be to firing rate. For our purposes here, whatever τ maximizes the linear fit of the function from frequency to burst rate should be chosen. *B* is ordered by $>_B$, the empirical greater-burst-rate relation. The first mapping function was introduced above, with $r_1 = A \rightarrow B$:

$$r_1(x) = x.$$

The second physiological relational system will define neural activity in subpopulation-1 of S1 which, recall, does not correspond to peripheral frequency either in terms of burst rate or firing rate, but rather in its temporal structure. In this case, again let $\mathfrak{B}$=the physiological relational system. To define $\mathfrak{B}$, we'll define the members of *B* in terms of PSFP, or power spectrum frequency at peak (Salinas et al. 2000). Briefly, recall that PSFP is calculated with a Fourier decomposition of the time course of neural activity, then the frequency bin with the peak power is found, and its median taken. This is the frequency that contributes most to the oscillatory activity of the particular neuron under consideration. Each member of *B* is a *frequency*, and so the natural ordering relation is the greater-frequency-than relation, $>_B$. Like r_1, r_2 is exceedingly simple, with r_2: $A \rightarrow B$:

$$r_2(x) = x.$$

Note that $r_1$is distinct from r_2: The first is a function from frequencies to burst rates, while the second is a function from frequencies to PSFP. Furthermore, PSFP is not a measurement of "more or less" periodicity, in the way that firing rate is a measure of how many spikes fire per second. It is rather a measurement of which frequency component of the overall activity of the neuron contributes most to its oscillatory activity. The final two functions I'll define describe the relationship between firing rate in subpopulation-2 of S1 and frequency, and then the firing rate of neurons farther downstream with negative slopes, relative to frequency. In each

[7] A function is bijective if it is *injective* and *surjective*. A function is injective (or one-one) if each member of the range is mapped to by only one element of the domain. A function is surjective (or onto) if every member of the range is mapped to by some element of the domain.

[8] More specifically, $\mathfrak{A}$ and $\mathfrak{B}$ are isomorphic if there exists a bijective function f: $A \rightarrow B$ such that for every *a* and *b* in *A*,

$$aRb \text{ iff } f(a)Sf(b).$$

If *f* is surjective but not injective, then $\mathfrak{A}$ and $\mathfrak{B}$ are *homomorphic*. A variety of other kinds of structure-preserving mappings can also be defined, by selectively loosening certain criteria. See (Swoyer 1991) for some examples.

case, the domain of B now consists of firing rates, and it is ordered by $>_B$, the greater-firing-rate relation. Let r_3: $A \rightarrow B$:

$$r_3(s) = 22 + 0.7s,$$

where s is stimulus frequency and $r_3(s)$ is rate described as a function of frequency. As reported in Salinas et al. (2000, 5506), this equation describes the relation between firing rate in S1 and stimulus frequency. (The equation also includes a noise term, but since noise is by definition not a signal, I've deleted the final term. Nonetheless, noise is a significant issue to be addressed; on this, see fn. 13.) Neurons in this population fire at a baseline rate of 22 spikes/s and increase linearly with a slope of 0.7 as vibration frequency increases. Finally, there are populations of neurons in S2 and beyond, which are oppositely tuned, whereby increasing frequencies generate decreasing firing rates (Salinas et al. 2000; Hernandez et al. 2000). To my knowledge, the specific equations describing the relations between the negatively sloped subpopulations and vibration frequency have not been published, though they are noted to be monotonic linearly decreasing functions.[9] For concreteness then, I'll stipulate r_4: $A \rightarrow B$ as

$$r_4(s) = 65 - 0.5s.$$

Although stipulated, r_4 should be considered as the equation that describes the activity of neurons in a population (either in S2, PFC, VPC, or MPC) with a negative slope relative to stimulus frequency.

Each of these four equations is an empirically discovered correspondence relation (with the exception of $r_4$which is stipulated; I'll omit that qualification henceforth) between neurons in specific populations and mechanical stimulation of the fingertip. These are the "specific correspondence relations" I've appealed to above in determining the teleofunctions of the neurons. Furthermore, the equations each define bijective functions that in turn define an isomorphism between the stimulus relational system $\mathfrak{A}$ and their respective physiological relational systems $\mathfrak{B}$.[10] The key idea here is that we find *systems* of representations, and *systems* of properties

[9] Furthermore, note that r_3 only describes the specific relationship discovered among neurons in subpopulation-1 of S1 with vibration frequency. Presumably, the populations of neurons in S2, PFC, VPC, and MPC, which also show positively sloped response profiles, admit of different specific relationships with stimulus frequency (i.e., different baselines and different slopes). They have not however been published (to my knowledge). Note that these different equations don't change the overall philosophical analysis of biological representation presented here; the theory easily accommodates differing correspondence relations between neural states and represented states, due to the versatility of the concept of structural preservation.

[10] Proving isomorphism is not trivial, and furthermore, measurement theory is concerned with one empirical and one numerical relational system, not two empirical relational systems as I've described here. But the technical details are outside the scope of this chapter, so I've made simplifying assumptions. Namely, I'll assume that $\mathfrak{A}$ and $\mathfrak{B}$ both have uncountable domains with countable order dense subsets, and their respective relations generate a total order on the domains. This suffices for isomorphism between two empirical relational systems $\mathfrak{A}$ and $\mathfrak{B}$ (Collins 2010, 406). Whether these assumptions are justified depends on whether making idealizing assumptions in general are justified.

represented, each organized in such a way that individual members from each domain map to members in the other, mapping specific firing patterns to specific vibration frequencies. I'll refer to these four functions as *representation functions*.

To determine representational content, recall that both causal etiology and structural preservation (e.g., isomorphism) are required. In each of the sensory representations throughout the vibrotactile discrimination circuit, the causal antecedent of the particular pattern of firing is the experimental stimulator. Thus, the thing to which each representation refers, determined by causal etiology, is the stimulator.[11] But causal etiology alone is not enough to determine predication, that is, to determine what property the representation predicates of the stimulator. For this, the representation functions for each respective neural population define which property is predicated of the stimulator and, crucially, determine which neural patterns would constitute accurate representation, and which would constitute error.

For example, assume that primary afferents in the rapidly adapting circuit are firing at a burst rate of 50 bursts/s and that this was caused by the stimulator. From r_1, we see that the representation function matches up frequencies to burst rate one-to-one; therefore, the representational content of this activity is something like *the stimulator is vibrating at 50 Hz*.[12] If the stimulator is indeed vibrating at 50 Hz, then the representation is accurate; if the stimulator is not vibrating at that speed, then the representation is inaccurate. But for neurons in subpopulation-2 of S1, where neurons have the teleofunction of corresponding to such external stimuli in terms of their firing rate rather than burst rate, and according to a different

[11] There are a variety of intermediate events between the stimulator's vibrating and a particular pattern of neural firing that it caused, say, in S2. For example, ion channels have opened and closed, neurotransmitters have been released, a variety of firing patterns have occurred in upstream areas in the spinal cord, brainstem, thalamus, internal capsule, S1, and so on. Determining which of these causal antecedents is the one to which the representation refers is known as the *causal chain problem*, which is a problem for any theory of representation that appeals to causation. While I won't attempt detailed discussion here, a reasonable solution (at least in this instance) is to appeal to teleofunction. The correlation of neural activity in S2 with upstream neural activity is not what confers survival advantage. Rather, by covarying with energy states at the periphery of the organism, in well-defined ways, distinct neural mechanisms can use that activity to perform transformations and computations which ultimately result in behavior that is appropriate to the environment. Hence, it is not arbitrary to claim that the neural activity refers to the stimulator and not some other link in the causal chain.

[12] Notice I write that the content is *something like* … (rather than that the content *is* …). It is unjustified to assume that the representational content of the lowest-level biological representations instantiated in the firings of individual neurons can be translated straightforwardly into a natural language. Rather, we should be satisfied with *describing* the content using natural languages, though should not expect a straightforward translation. Furthermore, note that it is equally justified to describe the content as "*that thing* is vibrating at…" as compared with "*the stimulator* is vibrating at…." The neural activity under question does not predicate the property of being a stimulator, only the property of vibrating at a certain frequency. Again, for the purpose of describing the content, rather than expressing or translating it, either rendering is acceptable because both expressions refer to the stimulator in this context.

representation function (r_3), if these neurons fire at a firing rate of 50 spikes/s, it does not imply that they have the same representational content. Rather, a neuron from subpopulation-2 of S1, whose teleofunction is to accord with external stimuli according to r_3, would, if firing at 50 spikes/s, have the representational content that *the stimulator is vibrating at 40 Hz* because r_3 maps the property of vibrating at 40 Hz to the firing rate of 50 spikes/s. If the stimulator is not vibrating at 40 Hz, then the representation is inaccurate. Similarly, a neuron that is part of an oppositely tuned subpopulation, say, in PFC, which has the teleofunction of corresponding to external stimuli according to r_4, would have a different representational content. Assuming again that it was firing at 50 spikes/s, this neural activity would have the content that *the stimulator is vibrating at 30 Hz* because this is the property that r_4 maps to 50 spikes/s firing rate. Similar comments apply to the temporal codes that use periodicity in S1.

In general, although the monkeys are quite good at the task (with about a 90% accuracy rate), they do occasionally make behavioral errors. When this occurs, there is a correlation between standardized measures of firing rate in S1 and S2 with behavioral error (Salinas et al. 2000). For example, if the monkey presses the lateral button, signaling that it believed that the comparison was *lower* when in fact it was higher than the base, the firing rates of its neurons in S1 and S2 are *less* than they would have been, had the animal made an accurate discrimination and *mutatis mutandis* for the opposite mistake. For example, assume that the comparison frequency is 40 Hz and that the base frequency was lower at 30 Hz. Since neurons in subpopulation-2 of S1 have the teleofunction of corresponding to superficial vibration pulses in their respective receptive fields according to r_3, in order to correctly represent the comparison stimulus of 40 Hz, the neurons should be firing at 50 spikes/s. Assume however that a neuron is firing at 40 spikes/s in this circumstance; in this case, its representational content is something like *the stimulator is vibrating at 25.7 Hz*, thus misrepresenting the frequency of the stimulator, which then leads, ultimately, to a behavioral error. In other words, sometimes a well-trained animal makes a mistake, signaling that it believes the comparison was lower on a trial in which the comparison was in fact higher. When this occurs, the neural firing patterns in early sensory areas (S1 and S2) fire at a rate that is lower than what it would have been, had the neurons accurately represented the stimulus frequency.

It thus appears that the behavioral error is a result, at least partially, of an early stimulus encoding error, where the sensory representations misrepresent the frequency of the stimulus. If only one or two neurons misrepresent that frequency, the animal's behavior as a whole will likely be unaffected. But as the number of neurons in error begins to mount, it becomes increasingly likely that the animal will behaviorally signal in error. Crucially, without accounting for the logical structure inherent in the representational content of neural activities, there is no way to make sense of the idea that the early sensory encoding mechanisms had *misrepresented* the stimulus, that is, that there was a stimulus encoding *error*. By accounting for both components of representational content, however, the struc-

tural preservation theory provides a theoretical framework that allows for such an interpretation.[13]

Structural preservation theory also applies to the sigmoid response profiles in motor cortex, which constitute generalized motor plans to press either the medial or lateral push buttons. These generalized plans become refined downstream by neural mechanisms in the basal ganglia, cerebellum, spinal cord, and motor neurons at the periphery. As with the sensory representations discussed above, we begin with the question of whether the neural activities in M1 are representations (at all), before addressing their content.

The behavioral output of pressing the medial versus lateral button in response to a comparison of two vibrating stimuli is learned, not evolved. Nonetheless, the animals do achieve high accuracy levels, and a reasonable teleological argument can be made on these grounds: The monkeys have learned that pressing the medial button when and only when the comparison stimulus is higher results in the acquisition of juice, and *mutatis mutandis* for the lateral button. Further, after learning, certain neural activities have come to be regularly correlated with the muscular motions associated with medial and lateral button-pressing. It is reasonable to conclude that the consumers of the neural activity in M1 (i.e., the neural mechanisms downstream of M1 in the basal ganglia, cerebellum, spinal cord, and motor neurons at the periphery) have the teleofunction of producing the state of affairs corresponding to the motor plan in M1. Or in other words, if the motor plan says something like *my right arm is pushing the medial button*, then the consumers of that motor plan have the teleofunction to make that true. This is analogous to my intention to pick up the coffee cup, which can be either satisfied or not. Thus, unlike sensory representations, whose teleofunction is to correspond to energy impinging on the periphery so that doing so is adaptive for the organism, the teleofunction of procedural representations or motor plans is to play a role in *bringing about* the states to which they correspond. In this case, the "direction of fit" is the reverse: Sensory representations "fit" the world; motor representations make the world "fit" them (cf. Searle 1992).

[13] As mentioned in the text above, the equation published in Salinas et al. (2000) includes a noise term, so should be written as: $r(s) = 22 + 0.7s + \sigma\epsilon$, where ϵ is noise with zero mean and unit variance and σ is the standard deviation of the mean firing rate. Since noise is by definition not a signal, I've deleted the final noise term. Nonetheless, noise in neural systems is a significant conceptual and practical issue to be addressed by a theory of representation; any plausible view must be able to account for it because there is no such thing as a noiseless signal in the brain. Many biochemical mechanisms such as ion channel opening, vesicle release, and ion diffusion are stochastic processes, so there will always be "random" electrical activity which is not a result of stimulus representation or neural computation. Although I don't have space for an in-depth discussion of this here, the theory on offer does have the resources to account for noise in neural systems. The general idea is to distinguish those alterations in the content-bearing properties of a vehicle of representation (e.g., firing rate) which are due to alterations at the source (e.g., vibrotactile frequency) from those alterations which are not due to alterations at the source; these latter alterations constitute noise. A firing rate that is within the range of noise, given its particular (empirically discoverable) noise range, representation function, and the value of its represented parameter, is a noisy-but-true signal, whereas one that is outside the noise range is a noisy-and-false signal. For more detail see Collins (2010, 359–363).

Recall that at the end of the comparison period, neurons in M1 correlate with neither the base nor comparison frequencies, but rather instead correlate with the difference, $f_2 - f_1$. Furthermore, there are again subpopulations with affinities for $f_2 > f_1$ and $f_1 > f_2$, respectively. Consider, for example, a positively sloped subpopulation (i.e., which "prefers" $f_2 > f_1$). As above, the specific equations defining the relationship between firing rate and $f_2 - f_1$ have not been published, to my knowledge, so I stipulate one for concreteness (and define a linear rather than sigmoid function for simplicity, but the conceptual points do not change). Notice that $a_1 = -a_2$, and that the constant is the point at which the function crosses the *y*-axis. Thus, if $f_2 = f_1$, the neuron will fire at the constant rate, and as f_2, the comparison stimulus, gets increasingly greater than the base, the firing rate increases as well.

$$g_1(f_1, f_2) = -2f_1 + 2f_2 + 44.$$

Notice that in this subpopulation, 44 spikes/s is the baseline rate, which increases or decreases depending on whether and by how much the base and comparison stimuli differ from each other. Unlike the sensory case however, these generalized motor plans only map to two outcomes: pressing the medial or lateral buttons. Thus, the mapping function from the set of firing rates to the set of behavioral outcomes very simply maps every firing rate from 0 to 44 spikes/s to something like *is pushing the lateral button*, and all rates above 44 spikes/s to something like *is pushing the medial button*. Note that this does not define an isomorphism between relational systems. It does however counter-preserve (but does not preserve)[14] the greater-firing-rate relation in the relational system composed of the two behavioral outcomes related very simply by the ordered pair, $\langle M, L \rangle$ (with *M* abbreviating "is pushing the medial button" and *L* abbreviating "is pushing the lateral button"). Thus, this mapping function fits within the broader construct of structural preservation and is an analogue of the technically more restrictive isomorphism.

Assume that a neuron in this subpopulation is firing at 55 spikes/s. Since g_1 maps this rate to the property *is pressing the medial button*, it follows that this neural activity predicates the property of pressing the medial button, of whatever it refers to. However, as before, reference is determined by causal history. Rather than referring to what caused them, however, procedural representations refer to what they caused. This reflects the reversed "direction of fit" of motor plans relative to sensory representations. Since the neural activity in M1 currently under consideration causes

[14] A function *preserves* a relation *R* only if $aRb \rightarrow f(a)Sf(b)$. A function *counter-preserves R* only if $f(a)Sf(b) \rightarrow aRb$, and thus, a function *respects R* only if it preserves and counter-preserves *R*; for isomorphism between relational systems, the mapping function needs to respect the relation *R*. As I mentioned earlier, there are good reasons to relax the strict requirements on isomorphism when using this tool to construct a theory of representation while keeping the basic idea of the preservation of internal relational structure across systems. The type of structural preservation appealed to in the text is a Δ/Ψ-morphism (Swoyer 1991), which preserves a subset of relations in one system while counter-preserving a subset of relations in the other (in this case, identity is preserved, while greater-firing-rate is counter-preserved; see Collins 2010, 329–330 for the details).

changes in the contraction levels of the various muscle groups of the animal's right arm, it follows that the representation refers to the animal's right arm. Hence, the representational content is something like, "my right arm is pressing the medial button." As with sensory representations, motor plans are semantically evaluable in the sense that they are satisfaction-evaluable; they can be satisfied or not. If the animal does in fact press the medial button, then the motor plan has been carried out; if not, then the motor plan or intention remains unsatisfied. This is the analogue of an inaccurate sensory representation. Note, as above, that different aspects of the representation determine different aspects of its content. Its bearing certain correspondence relations to behavioral outcomes, and having the teleofunction of producing the outcomes to which they correspond, makes them representations. The different firing rates are part of an ordered system, which correspond to a set of behavioral outcomes which also form a (very simple) ordered system, and the rates match up to the behavioral outcomes to which they correspond, determining an analogue of predication. Finally, causal etiology determines that the property of pressing the medial button is to be realized by the right arm.

The analysis of motor representations in monkey M1 is given at a far more abstract level than, say, the five pairs of motor neurons in the chemotaxis circuit of *C. elegans* discussed previously. In the latter case, the voltages of the motor neurons bear a continuous and specific relationship of proportionality to the degree of extension of muscles in the neck, which determine the neck's turning angle (and hence the direction in which the worm moves). This is due to the relative complexity of the different nervous systems (*C. elegans* has only 302 neurons). However, as the monkey's neural signals travel down the motor circuit and get closer to the periphery, the analysis of the content of motor representations will get more specific, analogous to the specificity of the sensory representations in early sensory processing areas. I consider this result – that structural preservation theory would analyze the neural activity in M1 in terms of abstract, generalized motor plans – to speak in favor of the theory. As I mentioned earlier, structural preservation is a versatile conceptual tool, and anything can be a member of a relational system, including relatively abstractly described behavioral outcomes.

6 Conclusion

The concept of representation, or at least *aboutness*, is the foundation upon which all other concepts of mental states and processes are built. To understand the place of mind in nature, we must understand what representation is and how living biological systems realize it. In this chapter, I have presented a sketch of a theory of biological representation and have illustrated it by appealing to the neurophysiological mechanisms involved in a sensory discrimination task. There are a variety of open questions that must be dealt with, including noise in neural systems and the causal chain problem. My main purpose for this chapter however was to outline and illustrate a *theoretical framework* that I think might be useful for making progress

on a theory of representation in biological systems. Whether that framework can support the detailed conceptual analysis required of a philosophically viable theory remains to be seen.

References

Akins, K. (1996). Of sensory systems and the "aboutness" of mental states. *The Journal of Philosophy, 93*(7), 337–372.

Bargmann, C. I., & Horvitz, H. R. (1991). Chemosensory neurons with overlapping functions direct chemotaxis to multiple chemicals in *C. elegans*. *Neuron, 7*(5), 729–742.

Bechtel, W. (2001). Representation: From neural systems to cognitive systems. In W. Bechtel, P. Mandik, J. Mundale, & R. S. Stufflebeam (Eds.), *Philosophy and the neurosciences: A reader*. Malden, MA: Blackwell Publishers.

Collins, M. (2010). *The nature and implementation of representation in biological systems*. PhD dissertation, Department of Philosophy, CUNY Graduate Center, New York.

Devitt, M., & Sterelny, K. (1999). *Language and reality: An introduction to the philosophy of language* (2nd ed.). Cambridge, MA: MIT Press.

Dretske, F. I. (1981). *Knowledge and the flow of information* (1st MIT Press ed.). Cambridge, MA: MIT Press.

Dretske, F. I. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.

Ferree, T. C., & Lockery, S. R. (1999). Computational rules for chemotaxis in the nematode *C. elegans*. *Journal of Computational Neuroscience, 6*(3), 263–277.

Fodor, J. A. (1975). *The language of thought*. New York: Crowell.

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.

Fodor, J. A. (1990). *A theory of content and other essays*. Cambridge, MA: MIT Press.

Fodor, J. A. (2008). *LOT 2: The language of thought revisited*. Oxford/New York: Clarendon Press/Oxford University Press.

Gardner, E. P., & Kandel, E. R. (2000). Touch. In E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), *Principles of neural science*. New York: McGraw-Hill.

Gardner, E. P., Martin, J. H., & Jessell, T. M. (2000). The bodily senses. In E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), *Principles of neural science*. New York: McGraw-Hill.

Hernandez, A., Zainos, A., & Romo, R. (2000). Neuronal correlates of sensory discrimination in the somatosensory cortex. *Proceedings of the National Academy of Sciences USA, 97*(11), 6191–6196.

Hernandez, A., Zainos, A., & Romo, R. (2002). Temporal evolution of a decision-making process in the medial premotor cortex. *Neuron, 33*(6), 959–972.

LaMotte, R. H., & Mountcastle, V. B. (1975). The capacities of humans and monkeys to discriminate between vibratory stimuli of different frequency and amplitude: A correlation between neural events and psychological measurements. *Journal of Neurophysiology, 38*, 539–559.

Mandik, P., Collins, M., & Vereschagin, A. (2007). Evolving artificial minds and brains. In A. C. Schalley & D. Khlentzos (Eds.), *Mental states, Vol. 1: Nature, function, evolution*. Philadelphia: John Benjamins Publishing Company.

Millikan, R. G. (1984). *Language, thought, and other biological categories: New foundations for realism*. Cambridge, MA: MIT Press.

Millikan, R. G. (1989). Biosemantics. *The Journal of Philosophy, 86*(6), 281–297.

Millikan, R. G. (2004). *Varieties of meaning, the Jean Nicod lectures*. Cambridge, MA: MIT Press.

Mountcastle, V. B., Talbot, W. H., Sakata, H., & Hyvarinen, J. (1969). Cortical neuronal mechanisms in flutter-vibration studied in unanesthetized monkeys: Neuronal periodicity and frequency discrimination. *Journal of Neurophysiology, 32*, 452–484.

Mountcastle, V. B., Steinmetz, M. A., & Romo, R. (1990). Frequency discrimination in the sense of flutter: Psychophysical measurements correlated with postcentral events in behaving monkeys. *The Journal of Neuroscience, 10*, 3032–3044.

Nicolelis, M., & Ribeiro, S. (2006). Seeking the neural code. *Scientific American, 295*(6), 70–77.

Romo, R., & Salinas, E. (2001). Touch and go: Decision-making mechanisms in somatosensation. *Annual Review of Neuroscience, 24*, 107–137.

Romo, R., Brody, C. D., Hernandez, A., & Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature, 399*, 470–473.

Romo, R., Hernandez, A., Zainos, A., Lemus, L., & Brody, C. D. (2002). Neuronal correlates of decision-making in secondary somatosensory cortex. *Nature Neuroscience, 5*(11), 1217–1225.

Romo, R., DeLafuente, V., & Hernandez, A. (2004a). Somatosensory discrimination: Neural coding and decision-making mechanisms. In M. Gazzaniga (Ed.), *The cognitive neurosciences*. Cambridge, MA: A Bradford Book. MIT Press.

Romo, R., Hernandez, A., & Zainos, A. (2004b). Neuronal correlates of a perceptual decision in ventral premotor cortex. *Neuron, 41*(1), 165–173.

Salinas, E., & Romo, R. (1998). Conversion of sensory signals into motor commands in primary motor cortex. *The Journal of Neuroscience, 18*(1), 499–511.

Salinas, E., Hernandez, A., Zainos, A., Lemus, L., & Romo, R. (1998). Cortical recording of sensory stimuli during somatosensory discrimination. *Society for Neuroscience Abstracts, 24*, 1126.

Salinas, E., Hernandez, A., Zainos, A., & Romo, R. (2000). Periodicity and firing rate as candidate neural codes for the frequency of vibrotactile stimuli. *The Journal of Neuroscience, 20*(14), 5503–5515.

Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.

Swan, L. S., & Goldberg, L. J. (2010). How is meaning grounded in the organism? *Biosemiotics, 3*(2), 131–146.

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

Vallbo, A. B. (1995). Single-afferent neurons and somatic sensation in humans. In M. Gazzaniga (Ed.), *The cognitive neurosciences*. Cambridge, MA: A Bradford Book. MIT Press.