Neuroscience and teleosemantics

Ruth Garrett Millikan

Synthese
An International Journal for
Epistemology, Methodology and
Philosophy of Science

ISSN 0039-7857

Synthese
DOI 10.1007/s11229-020-02893-9

Springer

Synthese
https://doi.org/10.1007/s11229-020-02893-9

INTERDISCIPLINARY PERSPECTIVES ON LOCATING REPRESENTATIONS IN BRAIN

Neuroscience and teleosemantics

Ruth Garrett Millikan[1]

Received: 15 July 2020 / Accepted: 25 September 2020

Abstract

Correctly understood, teleosemantics is the claim that "representation" is a function term. Things are called "representations" if they have a certain kind of function or telos and perform it in a certain kind of way. This claim is supported with a discussion and proposals about the function of a representation and of how representations perform that function. These proposals have been retrieved by putting together current descriptions from the literature on neural representations with earlier explorations of the features common to most things we are inclined to call representations (… maps, graphs, human language, signals between animals, stop signs … etc.) as these were assessed in Millikan (Language, thought and other biological categories. MIT Press, Cambridge, 1984 and following). Of interest is the degree to which these independent sources converge. I conclude that there is no need to employ any new or technical sense of the term "representation" for it to play an important role in neuroscience.

Keywords Neural representations · Teleosemantics · Representations · Philosophy and neuroscience

The call is for papers that will help us to "determine" "what 'representation' means" so that neuroscientists won't talk past one another when using this term. Clearly this is not a call for lexicography. Nor is it a call for conceptual analysis, for criteria we would supposedly use in sorting out things in any possible world. What may be useful is to ask whether there are properties common and peculiar to all or to a dominant portion of the common things *we actually* call "representations" and if so whether things of this kind might be found in the brain so that looking for them might be helpful for the progress of neuroscience. Is there anything interesting that is common to indicative sentences of many kinds, along with maps, charts, diagrams, graphs, road signs, bee dances, mating dances, animals' danger signals, labels on bottles and so forth that

✉ Ruth Garrett Millikan
ruth.millikan@uconn.edu

[1] Philosophy Department, University of Connecticut, Storrs, CT, USA

Published online: 26 October 2020

makes it natural to call all of them "representations"? If so, the hope would be that no special or technical meaning for "representation" needs to be fashioned for it to be naturally and fruitfully used in neuroscience.

In (Millikan 1984) and several works following I offered support for the proposal that there are properties common and peculiar to most of the ordinary things *we* call "representations." Here I want to compare results argued for in those earlier studies with contemporary views on what a representation in the brain would be. These views on what does or should count as a representation, derived in quite different ways at different times, show a surprising convergence. I offer this convergence as a kind of evidence for their soundness.

"Representation" is primarily a "function term," in a sense best explained by example. Suppose that in preparation for shipping a pile of scrap metal, the crunching machine happens to crunch various bits of this and that together so as to form an object precisely like a 1940s Underwood typewriter, but one that has been run over by a truck. Beside it on the scrap pile rests another object identical in every way that actually was formed by a truck running over a 1940s Underwood typewriter. That scrap is a badly smashed typewriter. It is a typewriter, but it is smashed. The other object, though identical, is not a typewriter at all. It is only exactly like a badly smashed typewriter.

> This analysis accords, at least, with the dictionary which says that a typewriter is a machine with keys for producing alphabetical characters, numerals, and typographical symbols one at a time on paper inserted around a roller. (American Heritage Online Dictionary)

It's not a machine with keys *that does produce* alphabetical characters nor is it a machine with keys that *could* produce alphabetical characters. It is a machine with keys ***for*** *producing* alphabetical characters.

The term "typewriter" is used for things that have keys "*for*" producing alphabetic characters, whether or not they are actually capable of doing this. That is what makes "typewriter" a function term, a thing's "function" being its purpose, what it is *for*. Whether or not a thing has a function and what that function is makes this difference: it determines whether it can be evaluated as a good or excellent or correct one rather than as defective, poor, not working or incorrect. It determines a standard against which it is to be measured, whether and how well it can perform this function. Unless a thing is taken to have a function, a purpose, a telos, there is no standard for this kind of evaluation.

Enter teleosemantics, with the claim that "representation" is a function term. The evidence? If it were not there would be no standard by reference to which representations would be correct or incorrect, true or false, fulfilled or unfulfilled, satisfied or unsatisfied.

> If language device tokens and mental intentional states (believing that, intending to, hoping that) are members of … function … categories, then they are language devices or intentional states not by virtue of their powers but by virtue of what they are

> supposed to be able to do yet perhaps cannot do. For example, just as hearts and kidneys are sometimes diseased or malformed, so sentences and beliefs are sometimes false, and words and concepts are sometimes ambiguous and sometimes vacuous. Such sentences, beliefs, words, and concepts are not able to perform their … functions. (Millikan 1984, p. 17)

That is all there is to the claim that was later dubbed "teleosemantics."

Teleosemantics says nothing about what the function of a representation is or about how a representation carries out that function. Despite widespread confusion on this matter it says nothing about how the "content" of a representation is determined. It leaves almost all of the work undone. Also, it is, or was originally, a claim about representations generally, not merely mental representations.

Once clearly articulated and correctly understood, the teleosemantic move seems perfectly harmless and indeed, hardly disputable. Consider the reflection of trees in calm waters which can yield all kinds of information about those trees yet when ripples distort that information, the reflection does not misrepresent. Simply as a reflection, it has no function, so it cannot fail to perform that function. Contrast conventional representations such as linguistic representations, artifacts with conventional symbiotic functions, paradigmatically, of passing information from speakers to hearers. If not formed in the right way they will be unable to perform this function but instead will misrepresent. If there are mental representations, they have a function or purpose at least in the sense that the heart has a function or purpose, so they too must be capable of misrepresenting.

Now if understood correctly, the teleosemantic move *should* have neatly divided the ongoing quest for representations into two completely independent parts. First there is the question what it is to have a function in the appropriate sense. What determines that this or that kind of thing has a function and the function that it has? Specifically, what determines the function, the telos, of this or that sort of representation? How does it acquire that function?

The second question divides in two: (a) what is the specific function of a representation as such, the function required for it to be a representation? What is its job? (b) What is the particular method, if any, by which a good undamaged representation performs its function? An undamaged typewriter produces alphabetic characters, numerals and so forth one by one by having keys that a user depresses one by one. Is there then a particular way, a particular mechanism, by which a representation serves its function?

The first question, question (a), was addressed. Most of those who examined or adopted the teleosemantic position assumed that a thing's function was what it had been designed or selected for doing, by a person, by natural selection or by some analogous developmental or social learning process. If there are representations in the brain they were produced by mechanisms selected by natural selection for producing representations. The body is filled, after all, with thousands of the most intricate and clever mechanisms that no Darwinian would doubt are there owing to natural selection.

But there was little or no attempt to develop a substantial answer to the second question, (b). No theories were offered about the defining functions of representations generally or about the ways they perform their functions. A number of (highly simplistic) competing theories concerning the nature of an inner or *mental* representation

were suggested and found wanting.[1] These experiments were often taken as tests of teleosemantics itself rather than as the tests of mini-theories of representational function that they actually were. There have been no arguments that I know of either for or against teleosemantics itself except the original argument that it must be right if there is to be such a thing as "Misrepresentation" (Dretske 1986).

How, then, should one go about answering the question "what is the function of a representation, *as* a representation?" The term "representation" might be hopelessly equivocal either at the center or around the edges. But we could discover this only by actually looking carefully to see if the bulk of things we would unreflectively call "representations" either do or don't, in fact, have a common function and whether they have a common way of performing that function. Done right, that would mean carefully comparing, say, maps and graphs and diagrams and informative sentences and road signs and certain kinds of gestures, railroad crossing signs, and perhaps animals' danger signals and mating dances. … Should we find a function and a way of performing it common to these, then we could go on to ask what evidence there might be of things in the brain that perform that kind of function in that way.

Trying to find commonness among the functions and ways of functioning of a diverse variety of familiar kinds of things we call "representations" was part of the project of (Millikan 1984, 2005, 2017).There each variety was shown to have a function in the same sense of "function" ("proper function"), to have, broadly speaking, the same outcome function and the same way of performing that function. Here I compare the results of those studies with influential contemporary views on representation in the brain.

If "representation" is primarily a function term, the leading question to be answered seems to be: what are representations for? What do they do when performing as designed? Like a typewriter or an axe, presumably representations are designed to be used. Thus the central claim of Ramsey's foundational book *Representations Reconsidered* (2007) was that an entity qualifies as a representation if and only if it is *used* as a representation.[2] What then is it for a representation to be used as a representation? Used by what for what? Ramsey did not answer those questions.

A very abstract answer to "used by what?" may not be that difficult to give, however. A representation is for use by some kind of interpreter, by something that uses it as a representation in performing some function that it has. This is a very slender beginning but it offers this much to the neuroscientist looking for representations. A representation in the brain would have to be used as such by some other part or aspect of the brain or by something connected to the brain. If there are representations in the brain there must also be interpreters for them. For the simplest sort of example,

[1] See, for example, the references given in the Stanford Encyclopedia of Philosophy under "Teleological Theories of Mental Content."

[2] An exception: unless they are intended as illustrations, intended to be instructive, depictions such as drawings are not defined as such by any particular use but they are called "representations". This sense of "representation" was one that Ramsey was especially concerned to exclude from talk of representations in the brain and it does seem a very different sense from that in which sentences and even, say, maps are representations. In any event, this sense of "representation" has been excluded from contemporary discussions of representations in the brain.

a reliable coincidence between certain neuronal activity in the visual system and something before the eyes that appears to be its cause is not yet evidence that the former is a representation of the latter. It would need to be shown that some part of the brain uses or reacts to this neuronal activity *as* a representation of that something's presence before the eyes. Which requires us to face the question what it is to use a representation as a representation in the performance of some function.

That if there are representations there must be interpreters for them need not be taken narrowly. Mental representation interpreters might be extremely complex systems involving large neural networks. Representation interpreters might use various mental representations together, say, in performing inferences. A representation in the brain might have many potential interpreters or interpreters that have many different settings, different uses, as one's perception of an empty beer can on the trail might guide one in stepping over it, picking it up with one's hands, picking it up with a stick, kicking it along the trail or into the woods, stepping on it, pointing to it, considering the character of the person who left it, and so forth. Supposing beliefs are mental representations, a single belief such as that animals can drown in water or that it's lighter in the day than at night can be used to advantage in different ways many hundreds of times during a lifetime.

What then is it for an interpreter to use a representation *as* a representation? (A typewriter might be used as a rather blunt weapon, but that would not be using it as a typewriter.) A pretty standard contemporary answer from cognitive scientists is that a neural representation would be something that "guides behavior," or that plays a role in "controlling actions," when performing tasks that require taking account of that which is represented.[3]

The behavior or action of what? The behavior or action of the whole animal is what seems to be meant and this, of course, is what the neuroscientist eventually wants to explain. But many things besides external behavior are controlled by the brain, many things that go on in the rest of the brain and in the body. Many things that go on in the brain itself must be synchronized. Feature detection helps to instigate deeper perception, inferences and other transitions of one mental state to others seem to occur. Surely the question is open whether some of this work is done using methods that work on the same principle as do methods of controlling external behaviors.

And it is easy enough to generalize. Every kind of representation,[4] including a neural representation, should there be such, is for guiding the behavior or operation of another mechanism or other mechanisms, "interpreters" or "consumers," having functions that can be performed successfully only if their activities are set or molded to fit or take account of the state of affairs represented. Representations are used by their interpreters for guidance in performing functions of the interpreter. If the representation

[3] For two of many examples, see (Piccinini 2018), Schulz (2018).

[4] Including fact-stating sentences, maps, diagrams, animal signals and so forth (Millikan 1984).

is correct, an interpreter can perform a certain task or tasks of its own by "using" or being adjusted by the representation to take account of the represented state of affairs.[5] That the representation is correct will then be part of a causal explanation of how the interpreter manages to perform that function.[6]

Finally, we should ask whether representations have a common way of functioning (as typewriters have keys to be struck). Here again there seems to be an emerging consensus, for example (Palmer 1978, Millikan 1984, Gallistel 1990, 2008; Swoyer 1991; Ramsey 2007; Piccinini 2018). Best known, perhaps is Gallistel's 1990 statement:

> a representation in psychology is a functioning isomorphism between processes within the brain or mind … and an aspect of the environment to which those processes adapt the animal's behavior. (Gallistel 1990, p. 2)

Ramsey called this "S-representation." In (Millikan 1984) it was just called "representation."

> There are operations upon or transformations (in the mathematical sense) of the [representation] that correspond one-to-one to operations upon or transformations of the [represented] such that … any transform of the [representation] resulting from one of these operations has as a Normal condition for proper performance the corresponding transform of the [represented]. (Millikan 1984, p. 107)

Representing requires an isomorphism (technically, a homomorphism) between the set of representations in a system and the set of things they represent such that an "interpreter" mechanism or mechanisms can use the representations to adapt its ways of functioning to that which is represented as is required to perform certain of its own functions. To be guided by a certain aspect or aspects of the form of the representation is equivalent to being guided by a certain aspect or aspects of the represented. Considering that the form of a representation—the values of its significant variables—is determined as a function of the form of its represented, it is not mysterious that the representation's form might help to guide an activity that requires taking account of the form of the represented. The use of representations is an engineering principle, like the use of levers or gears.[7]

---

5 … [a representation] serves to adapt the cooperating interpreter device to conditions such that proper functions of that device can be performed under those conditions (Millikan 1984, p. 98).

6 Unless, of course, the function is performed instead by some sort of lucky accident.

7 Current literature sometimes requires of mental representations that they are normally caused by their represenTeds.

> … there is an emerging consensus that the best way to understand representation in the context of cognitive explanation … includes four elements: (I) a homo-morphism (partial isomorphism) between a system of internal states and their target, (ii) a causal connection from the target to the internal states, … (Piccinini 2018, p. 3).

Certainly any producer of representations of facts has to have a method or methods, reliable or unreliable, of detecting those facts, but there are many kinds of natural information about things to be drawn from

A homomorphism holds between two *sets* (not two individuals). Every representation is, as such, one of a set of possible representations representing different alternatives, alternative possible states of affairs.[8] These form a network such that certain (mathematical) operations that would transform one representation in the system into another representation in that system are paired with operations that would transform what the first represents into what the second represents. It follows[9] that there is a rule or rules by which the representations in a system determine their representeds.

Like a sentence, then, a representation must be articulate. The articulation of a representation is determined by the kinds of transformations that would transform one representation in the system into another. Each feature of a representation that would produce a different representation if transformed, corresponds to a variable in the representational system. What remains unchanged under every meaningful transformation is the invariant of the representational system. Thus every representation must have at least two significant parts, an invariant and at least one variant.

This raises a question about certain simple representations, ones that are more naturally called "signals" or just "signs" than "representations." A flashing light at a railway crossing that tells when the train is approaching is called a "signal" not a "representation." An exit sign or crossroad sign on the highway is called a "sign," not a "representation." Whatever its "content," the excitation in certain optic nerve fibers that causes the frog's tongue to flip out after a (possible) fly might more naturally be called a "signal" than a "representation."[10] The reason is that these signs do not appear to be articulate. They do not appear to have any features that, if changed, would yield a different meaning. They do not appear to be part of a projectable representational system in which something else might have been said instead.

But this appearance is mistaken. The difficulty is that to discern the system in these simple cases requires adopting an unusual perspective, one from which time and/or place are seen not as mere context for these representations but as significant variables

Footnote 7 continued
besides information gathered from their effects. True and well-evidenced representations can be of future events, the direction of the North Star carries information about the direction of the North Pole, and so forth (Millikan 2017 §10.2).

[8] Representational systems can be very small. The system with lanterns (one if by land and two if by sea) that the Sexton of the Old North Church used to tell Paul Revere when and how the British were coming could have said "by land" or "by sea" and could have said it on any day the British began approaching. However many days one supposes Paul and the Sexton would have waited, two times that many days is the total number of possible messages that that system could have carried.

[9] Roughly, because if A is to B as C is to D then A is to C as B is to D.

[10] This particular example from Fodor (2008) has produced a complex literature. The excitation in the frog's optic nerve is produced by any small black particle that moves in certain ways across the frog's line of vision. The question has been raised whether the tongue-snapping reflex that is guided by the time of the optic nerve excitation could be said to represent presence at that time not of a fly but merely of a black particle. The truth of the representation would then figure in an explanation of the success of the tongue reflex as follows. Given where frogs hunt, it is statistically probable that any black particle moving in that certain way is a fly. So if the frog ingests a black particle it is statistically probable that it injests a fly. Further, the tongue reflex is not malfunctioning when it injests a black particle that is not a fly. So, it is argued, whether the optic nerve excitation means fly here now or black particle here now (or any of a number of other things) is indeterminate.

The major mistake made here is that the explanation is not a causal explanation of success as is required but a statistical explanation. A minor difficulty is the failure to distinguish between malfunction and failures that are owed to other conditions.

within them. Further, it requires interpreting these variables as self-signs, that is, as standing for themselves. Compare the color painted on the outside of a colored pencil that stands for itself as the color of the crayon inside. If one didn't know the time and/or place of a flashing rail-crossing light one would not know the time or place that a train was approaching. Clearly time and place carry some of the light's message. That the light at Jones crossing was flashing at 5:03 pm on 12/19/93 represented that a train was approaching Jones crossing at 5:03 pm on 12/19/93; that the light at Smith's crossing was flashing at 8:23 am on 12/19/93 represented that a train was approaching Smith's crossing at 8:23 am on 12/19/93. The system is simple, yet it contains an uncountable number of possible signs.[11]

An answer can be given then to some who have complained when neuroscientists take what seem to be mere signals employed in neural functioning to be representations.

> Representation" is commonly used in psychology and neuroscience in the sense of a mental or neural entity (for example, the activity of a feature detector) that is imagined simply to "stand for" some nonneural entity … Such a representation is nominal in both senses of the word. It functions only as a unique naming device. (Gallistel Gallistal 1989, p. 159).

Like complete sentences, representations make claims. They are not mere names. But Gallistel is mistaken about feature detectors, for they make claims about when and where a certain feature is to be found. Signals are a genuine kind of representation.

However what follows is, for instance, that adrenaline running in a person's blood stream at a time represents a need for vigorous action at that time, its interpreters being a widely dispersed collection of body mechanisms, including the brain, that respond, for example, by immediately increasing the heart rate and blood pressure, expanding the air passages of the lungs, and so forth. And so for each of myriad chemical signals sent from one part of the body to others. They work on the same principle as do representations. Refusing to call them "representations" would be failing to recognize an important generalization.[12]

Call signals "representations" or not, there seems to be no need to fashion any special or technical definition of the ordinary term "representation" for it to be fruitfully used in neuroscience.

References

Bechtel, W. (1998). Representations and cognitive explanations: Assessing the dynamicist's challenge in cognitive science. *Cognitive Science, 22,* 295–318.

Dretske, F. (1986). Misrepresentation. In R. Bogdan (Ed.), *Belief: Form, content and function* (pp. 17–36). New York: Oxford.

Fodor, J. A. (2008). Against darwinism. *Mind and Language, 23*(1), 1–24.

[11] For a discussion of the employment of self-signs in a variety of human communication systems see (Millikan 2017, Ch. 9).

[12] Following Bechtel 1998, the height of the arms of a working Watt governor do represent the pressure in the engine's cylander. There is nothing ethereal about representations. They are merely pieces or properties of things that are put to use in a mechanism that operates in accordance with a certain principle.

Gallistal, C. R. (1989). Animal cognition: The representation of space, time and number. *Annual Review of Psychology, 40,* 155–189.
Gallistel, C. R. (1990). Representations in animal cognition: An introduction. *Cognition, 37*(1–2), 1–22.
Gallistel, C. R. (2008). Learning and representation. In J. Byrne (Ed.), *Learning and memory: A comprehensive reference* (pp. 227–242). Amsterdam: Elsevier.
Millikan, R. (1984). *Language, thought and other biological categories*. Cambridge, MA: MIT Press.
Millikan, R. (2017). *Beyond concepts; unicepts, language and natural information*. Oxford: Oxford University Press.
Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 259–303). Hillsdale, NJ: Erlbaum.
Piccinini, G. (2018). Computation and representation in cognitive neuroscience. *Minds and Machines, 28,* 1–6.
Ramsey, W. M. (2007). *Representation reconsidered*. New York: Cambridge University Press.
Schulz, A. (2018). *Efficient cognition*. Cambridge, MA: MIT Press.
Swoyer, C. (1991). Structural representation and surrogative reasoning. *Synthese, 87,* 449–508.