

The Ontological Status of *Representations*

Matthias Scheutz

Department of Computer Science and Cognitive Science, Indiana University
Bloomington, IN47401, USA
email: mscheutz@indiana.edu

Abstract

The goal of this paper is to argue that the ontological status of *representations* can only be evaluated within a theory. In other words, what counts *as* representation, or whether a certain representation is *better* than another one, depends solely on the (level of) description of the phenomenon under scrutiny. It is shown how "representation", being a semantic notion, can be defined in terms of the notion "meaning". For cognitive science, in particular, it follows that representations, functioning as mere *descriptive* devices to facilitate one's goal of explaining and modeling brain/thought processes, cannot in and by themselves give rise to ontological or epistemological claims.

Introduction

Representation, as with all widely used terms, is a rather ambiguous notion that carries a lot of semantic overweight with it: in what sense does a word, for example, *represent* its meaning, a letter grade students' abilities, the retinal image objects, or nuclear missiles peace? It is obvious that one has to restrict all possible connotations of *representation* if one wants to discuss the applicability and usefulness of this notion for cognitive science.

Most cognitive scientists use "representation" to describe the relation of certain mind/brain¹ states to events in the world, i.e., the "inner" to the "outer" perspective, the assumption being that minds/brains have to somehow "represent" what is given in the world to make sense out of it, or as Gardner puts it:

*"[...] the cognitive scientist rests his discipline on the assumption that, for scientific purposes, human cognitive activity must be described in terms of symbols, schemas, images, ideas, and other forms of mental representation."*²

¹ At this point I prefer the somewhat imprecise "mind/brain" over a clear-cut distinction to allow for easier reference to both camps in cognitive science: the one which holds that brains use representations and the other which believes the same of the mind.

² See Gardner 1985, p. 38.

The term "represent" already suggests this usage, since "re"- "present", being composed of "re" and "present", literally means "present again".³ It indicates that something which is not *present*, but which had been *presented* at some point, is to be *presented again* (see also Glaserfeld 1995). What this reading means for cognitive science is that the brain exploits mechanisms to keep track of perceived objects and store this "impression" in a way that it can be retrieved and used if necessary. This "stored impression" is then labeled "representation of the stimulus". The need to use representations seems to arise from the lack of presence and/or persistence of stimuli in the perceptual system: what is not given in the perceivable environment anymore, but has been perceived at some previous time, must be recalled from memory when needed (not the object itself, of course, but the perception of the object at the time it was perceived). In short, one could claim that *representations* are required for things that are *not given perceptually*.⁴ In the following, however, I will argue that "representation" is a concept which exists only relative to one's level of description (=theory). Hence, the question "Does the brain *represent* the world?" as such cannot be meaningfully answered.⁵

³ Consulting the Webster Dictionary, one learns that "represent" stemming from the Latin "repraesentare" has at least five "meanings", three of which matter in particular to cognitive science: "represent" as "to present a picture, image, or likeness of", as "to describe having a specified character or quality", and as "to serve as a sign or symbol of <the flag represents our country>". The first two readings seem to correspond to what Harnad has called "nonsymbolic representations" (first corresponding to his "iconic representation", the second to "categorical representation"), whereas the third would describe his "symbolic representation" (see Harnad 1990 and Harnad 1993).

⁴ A more radical position is assumed by ecological psychologists who following Gibson do not believe in representations at all (see Gibson 1966). In fact, this topic is still vividly discussed between representatives of the "representational" view and Gibsonians (see, for example, Fodor and Pylyshyn 1981, and Turvey, Shaw, Reed, and Mace 1981).

⁵ Constructivists (e.g. Foerster) would confer the attribute "undecidable" upon this question.

Where is *Representation* Needed?

A major branch of cognitive science is devoted to describing the relation between brains (or, more generally, cognitive systems) and the "outer" world, i.e., events in the world and their "mappings" onto certain states of the brain. When cognitive scientists thus seek to explain the processes occurring and concurring in the human brain (or in the brain of any other living creature, for that matter) during perception, e.g., visual perception, they describe the flow of causation from the perceived object, to the "image of the object on the retina",⁶ to the patterns of neural activation, which is *caused* by this image, calling the last "the representation of the object (in the brain)".⁷ So the term "representation", being part of their descriptive vocabulary, is used to name/describe the result of the perceptual process, i.e., the result of *what* an individual perceives.⁸ "Representations" obviously come in very handy at this point, but not without a price to be paid for their convenience; their liberal application caused and causes fundamental ontological confusions. Just because the brain *can* be described as "holding representations of perceived objects" it does not follow that the brain *must* hold representations. On the contrary, the brain in and by itself might not need representations at all (see, for example, Sander 1996, or Llyod 1989)! Neurological research, for example, suggests that what we perceive consciously (the level of description of cognitive psychology) is very different from what actually happens at the neuronal level (the level of description of neuroscience). So at one level of description we seem to have representations of objects, whereas on another one we do not (I will address this issue again in the next section).⁹

⁶ Note that this "imaginary" language already suggests that there be a "representational mapping" between the object *per se* and the result of light cones bouncing off the surface of the object and hitting the retina.

⁷ Churchland is one of the representatives of the "patterns of activation as representation" camp (see Churchland 1989). Even Denett 1991, p. 191, switches to "representational" language.

⁸ Of course, this is not to say that what we subjectively perceive is a representation of that object, rather we (most of the times) have the *impression* that we perceive the object *itself* (we simply can't help it!). It is therefore already a (theoretical) ontological judgment, may it be merely an assumption or already the product of a reasoning process within given boundaries of a cognitive theory of perception, to hold that there is something "behind" our perception, the "real thing" or - in Kantian terminology - "das Ding an sich" (engl. the thing-in-itself), which eludes our perception. Hence the question whether or not we "perceive the object itself" does not matter during and for the actual perceptual process at all.

⁹ Interestingly enough, it is the neuroscientists and neuroscience oriented psychologists that are mainly "anti-

A more opaque philosophical problem arises from the common usage of the term "representation" (as a two-place relation between the thing to be represented and the thing being its representation), which tacitly forces ontological as well as epistemological commitment to the nature of those representations. The fallacy here is to neglect the theory-dependency of what counts as representation which immediately becomes apparent once "representation" is debunked as an inherently semantic concept (see also Llyod 1989, p. 9). For "being a representation of" or "being about" name essentially the same relation, the latter being better known under the name "meaning". Hence, *representation*, coinciding with (one reading of) *meaning*, crucially depends on one's theory of meaning.¹⁰ Among the various different analyses of the notion "meaning", a main distinction can be drawn between "meanings of" and "meanings for", the former being a two-place, the latter a three-place relation. Whereas the two-place version is basically a logical one, the three-place relation version always involves an individual: *x* means *y* for *z* (where *z* is a particular individual). Another way to characterize this distinction would be to use Haugeland's distinction between "derived" and "original" intentionality. Among the many reasons that make the latter very reading of *meaning* more appropriate for cognitive science than the two-place versions (e.g., the standard interpretation function from logic) is its account for the ordinary language phenomena such as the context- and/or speaker-dependency of the meaning of language expressions, extra-linguistic meaning, etc.

Taking this notion of "meaning" as primitive, a two-place version of "representation" can be extracted: *x* represents *y* if there exists a *z* such that *x* means *y* for *z*.¹¹ Note that *z* does not have to be confined to people, it could also stand for a theory (i.e., the theory of a person or a group of persons). By the same token, the predicate "symbol" can be extracted from either "representation" or "meaning": *x* is a symbol if there

representational" inclined whereas most computer scientists, logicians, and linguists believe in cognitivism (the computational claim on mind).

¹⁰ For the intended reading of "meaning" I refer the reader to "meaning #6" in Haugeland 1985, where also other interpretations can be found, since a thorough analysis of the meaning of "meaning" goes far beyond the scope of this paper...

¹¹ It seems that major parts of the semantic realm comprise interdependent, mutually definable, maybe even necessarily circular notions such "meaning", "reference", "denotation", "symbol", "representation", etc. which makes it difficult to discuss one without giving an account of the others, too (see, for example, McGee 1991 for a logical treatment of the interdefinability of certain semantic notions).

exist y and z such that that x means y for z , i.e. to *be a symbol* is to *be meaningful to somebody*.¹²

Even though "representation" is rendered two-place, it cannot and should not hide its three-place core which relativizes its ontological status: Representations do not exist independent of individuals for whom they are "meaningful". Take, for example, the visual system of a frog which has been extensively investigated. One of its salient features is the ability to recognize dark moving spots - the "bug-detector" (Letvin, Maturana, McCulloch, and Pitts 1988)! Consider now Kermit, the frog, seeing a fly landing on a blade of grass right in front of him. The black spot on Kermit's retina initiates activities in his brain eventually resulting in a tongue movement (in order to catch the fly). It has been argued that the dark spot on Kermit's retina be a *representation* of the fly. However, this could only be true for the theory of the observer, since the frog's brain itself does not have any conception of what a bug detector is; it simply "reacts" to a stimulus—or so it is claimed.¹³ Or take the circuitry of the computer I am working on right now which, for the specialist, "represents" the circuits of that very machine. This does not imply, though, that the shades and specks of color on the blueprint *actually* represent the computer's circuits.¹⁴ For a layman, it will just be a strange-looking collection of lines and circles without any semantic content. However, we would certainly not want to infer from this unfortunate fact that the diagram does not represent anything at all! And once representations are seen to be theory-dependent, it is quite easy to appreciate that everything can potentially be a representation of something, very much like every object can serve as a symbol.¹⁵

Levels of Description

In the rough analysis of object perception we have (implicitly) distinguished three levels of description: a subjective, a "representational", and a neuro/biological level. Whereas the so-called "representational" level

(of cognitive science)¹⁶ introduces the concept of "representation" to link objects in the world to objects in the mind, the two other levels do not involve representations at all: on the lower, the neuro/biological level, the concept of *object* is not even defined; on the higher, the subjective level, it is the appearance of the object we are aware of, its distance from our body, its persistence, its shape. In short, we are aware of the object *itself* and not of a *representation* (since what counts as an *object* is exactly what we are aware of as such).

In describing the nature of these three levels with respect to representation, we have tacitly ranked them according to some (intuitive) complexity measure. There is an obvious advantage to arranging sets of theories in hierarchies (i.e., theories that deal with the same subject area, but approach it with different methods, using different spatial and temporal "resolutions"): not only does this permit one to visualize possible dependencies of concepts across those levels, but it also enables researchers to try to reduce notions defined at a certain level of description to notions defined at what is considered a more "fundamental" level. Reductions are not only desirable because of their ontological parsimony, but they can also shed new light on the subject matter by establishing links between scientific approaches that aim at different goals.¹⁷ Since there is no designated "right" level,¹⁸ researchers are left to choose the (in their opinion) most appropriate level to explain the investigated phenomena. If a level turns out not to be well-suited for a given problem, it can always be abandoned in favor of a more explanatory one.¹⁹

¹² At first glance, this definition of symbol might seem somewhat *ad hoc*, but after more careful inspection it reveals itself as sufficient for most applications in cognitive science. Notice also that this definition is indifferent about the "usual" philosophical distinction between "symbols" and "signs", since it does not decide whether or not the meaning of x is fixed for z (x being a *sign* in the former, and a *symbol* in latter case in the sense of Langer 1948, for example).

¹³ See Llyod 1989, as well as Smith 1966, pp. 215 - 219, for a detailed discussion.

¹⁴ See also Putnam's ant example in Putnam 1981.

¹⁵ Clancey makes a similar point with respect to cognitive modeling when he suggests that "we reclassify most existing cognitive models as being *descriptive* and *relative to an observer's frame of reference*, not structure-function mechanisms internal to an agent that cause the observed behavior" (Clancey 1989, p. 108).

¹⁶ One has to be careful with the usage of "representational": Every theory *per se* is "representational", since it involves symbols. Also, theories talk *about* representations explicitly, making the notion "representation" part of their either primitive or definable vocabulary. What is meant here with "representational" level is the second, namely that there exists a level of description in cognitive science, regarding the subject "brain and cognition", which is situated above "neuroscience" and below "philosophy of mind", for example, and deals with "representations" explicitly (i.e., representations of objects in the world within the brain).

¹⁷ Establishing connections between hitherto unrelated theories has not seldom initiated an intellectual quantum leap, e.g., the link between quantum mechanics and the theory of relativity, or the one between logic and computability. Most cognitive scientists, I take it, are currently dreaming of a similarly revealing link between neurobiological concepts and the human consciousness.

¹⁸ It is crucial to this method that there are no "right" descriptions; all those descriptions are mere *constructions* of our conceptual system.

¹⁹ Obviously, much depends on one's theoretical stance with respect to scientific theories: instrumentalists will certainly be more generous than realists in this respect.

Although hierarchies of levels of description can be constructed for virtually any given phenomenon, the internal structure of those hierarchies depends crucially on the subject matter. Neither do hierarchies have to be well-founded (i.e., have a "bottom" level), nor linear (i.e., every level has exactly one directly below and one exactly above), nor even finite.²⁰ Their levels of description can generally be distinguished in terms of time and space intervals at/in which their "atomic" (i.e., at the given level irreducible) entities are defined.²¹ Some levels might be reducible to lower levels by showing how "higher level concepts" can be defined in terms of lower level ones.²² Take, for example, the mathematical construction of "equivalence relations". It allows one to "abstract away" from low level objects by forming "equivalence classes", a new entity with new properties which, nevertheless, can be defined in terms of collections of low level entities.²³ Others might not be reducible at all, causing insurmountable difficulties for a reductionist program.²⁴

Another problem worth investigating is which of the properties that hold on one level propagate upward or downward. Take, for example, "levels of implementations" in real computers where at some stipulated "bottom level" an assembly programming language is implemented. Then another programming language can be implemented on top of the assembly language, and so forth. Now it might be interesting to know whether properties such as "Turing computability", "real-time", "reference", etc., in particular "representation" cross implementation boundaries.²⁵ In fact, the very same problem is of essential importance in cognitive science when one

considers the various levels at which cognitive systems can be described. This goes to prove, once again, how valuable a model the computer has become for cognitive science even without subscribing to computationalism!

Cognitive science, consisting of various fields of individual sciences, naturally attempts to tackle cognitive phenomena from different angles depending on the research methods of the respective contributing discipline, thereby encouraging different levels of descriptions. Some of these levels are known to be reducible to others, but are still used for convenience and/or complexity reasons (e.g., the functionality of certain groups of neurons, so-called modules, viewed as a "black box"). Whether all levels are eventually reducible, or whether irreducible ones exist as well, is still an open problem. However, it is generally believed that what is described on upper levels is *caused* by what is described on lower levels.

This is where the "representation"-issue comes into play: whether or not something counts as representation of something else is dependent on the level of description! If this level is reducible to a lower level where what counts as representation on the upper is no longer representing anything, then Occam's "razor" can be used to eliminate superfluous entities from one's ontology. This, of course, is not to say that the level of description has therefore also become superfluous, since properties defined on the upper level are usually not definable at the lower one. Furthermore, they might offer more insight regarding the nature of the studied phenomenon than the ontologically "purer", reduced description does. To illustrate this point, consider pointer variables in a programming language, say PASCAL. By PASCAL's semantics, contents of pointer variables *denote* other variables; they "point" to them. But at the level of machine language, in which PASCAL is implemented, they are just *memory locations* (which do not distinguish between pointer or other variables).²⁶ However, the way those memory locations are treated by the assembly program suggests that at a higher level of description they could be viewed as "denoting" something. Another example for the theory/level-dependency of representations, once more taken from computer science's sheer inexhaustible richness of abstract entities, is the "representational status" of bits (=binary digits) as exemplified by the question whether or not bits are actually symbols. Following the above described reduction of symbols to representations, the question can be reformulated: are bits representations of some *y*? And again, the laconic answer is: it depends!

²⁰ For nonwellfoundedness consider the ongoing discussion in physics whether there are smallest particles, for linearity and infinity consider a digital hardware, on which a binary number system can be implemented. The binary system is then used to implement a ternary system, and so forth... Of course, for any given *n*, to implement the *n*-ary system, one does not have to implement all the others below, so the hierarchy is potentially infinite and not necessarily linear.

²¹ Notice that interdependence of space and time seems to permeate many different levels of description in cognitive science (e.g., in neuroscience it is the size of neuron and its firing rate, the size of modules and their computation time, see Newell 1990, and also Edelman 1992, p 124; in computability theory it is the intertwined space and time complexity classes).

²² Those levels of description might be called "levels of abstraction".

²³ An interesting question arises in this connection: Even though a level might be reducible to another level, this reduction need not be computable. What are the consequences?

²⁴ See Quine's example about "Vienna" in Quine 1987.

²⁵ For a detailed discussion about which properties cross implementation boundaries see Smith 1966, pp. 40 - 42.

²⁶ Although some assembly languages provide commands to implement pointers directly (such as "MOVE indirect"), once they got assigned a memory location by the compiler, they are indistinguishable from the contents of normal variables. In other words, the "signifier-signified" distinction has vanished.

In order to demonstrate this, we follow the classical method to debunk antinomies providing arguments for both possible answers. First, we will argue that bits are in fact symbols, since in some sense, bits (i.e., binary digits) can be viewed as models of certain states in a digital computer which at a lower level can be described as an analog machine (since the circuits all use certain voltages and the transition between those is continuous). To show this one needs to map "voltages" (i.e., real numbers which are used to model voltages) onto binary numbers. For example, one could define the following interpretation function f from the Reals into $\{0,1\}$:

$$f(x) = \begin{cases} 0 & \text{if } x < 3.0 \\ 1 & \text{if } x \geq 3.0 \end{cases}$$

Now consider a store element (e.g., a capacitor) to which a certain voltage is applied, say 2.4 Volts. Then f will assign "0" to that capacitor and therefore *model* its current state (although in a very "coarse" way since 2.3 and 2.5 Volts and so on will also only get mapped to "0"). It is now possible to build a formal model of the machine using the bits "0" and "1" and describing the functionality of the machine not in terms of potentials, capacities and currents, but in terms of changes of zeros and ones. This formal model is, of course, just an abstraction, it does not *exist* on the physical machine. It is possible now to define "bit" within this formal model:

$$\forall x (bi(x) \leftrightarrow x = \bar{0} \vee x = \bar{1})$$

Therefore "bits" are symbols.

Interestingly enough, the opposite route also works, taking a digital, say binary, machine and mapping it onto an analog hardware. Let f be the function from $\{0,1\}$ into the reals defined as $f(0)=.5$ and $f(1)=5.0$ (maybe 0 and 1 could also get mapped onto an interval of reals). Then the digital description of the machine can be implemented directly on the electronically described machine, provided that the crucial structural description are preserved under the mapping (i.e., certain predicates describing the functionality of the machine). In this case, the Reals are used to *model* bits, hence Reals are the symbols used to describe the digital (binary) machine, in other words, bits are taken to be the fundamental building blocks and are therefore not symbols. Another possibility would be to group 7 bits together to implement the standard ASCII code, where "bit" is taken to be the fundamental notion and "ASCII-character" to be definable in terms of bits (i.e., ASCII would then be the derived notion).

At this point we will depart from the classical argumentative structure since we do not want to conclude that "bit" is a paradoxical concept, but instead the above arguments should show that whether or not bits are representations solely depends upon the level of description, i.e., whether bits are taken to be original or derived entities.

Many more examples can be found in different disciplines, revealing the plethora of human symbolic potential, which nevertheless will always be confined within its own descriptive boundaries, summarized in the modified version of a dictum from Wittgenstein's *tractatus*: "Die Grenzen meiner Sprache sind die Grenzen meiner Theorien" (engl. the limits of my language are the limits of my theories).

Conclusion

Representations are, no doubt, extremely useful explanatory linguistic devices which not only allow us to talk "about" things, but are also *conditio sine qua non* for scientific theories. However, as I have argued above, being bound to theories (in particular, levels of description) their "existence" is inevitably relativized: what might count as representation in one theory might not in another (if it can be compared at all). As main consequence for cognitive science, the claim that a "representational level of description" is necessary to understand mind has to be abandoned unless it can be stringently shown that this level will provide irreducible insights, i.e., notions that cannot possibly be explained in terms of the vocabulary of lower levels of descriptions. Although various such attempts have been made throughout the intellectual history, none of them, to my knowledge, has born fruit so far. Given this fact combined with the arguable "arbitrariness" of representations, it seems very unlikely that convincing arguments for "irreducible concepts" are even possible in principle, let alone in sight. Hence, appreciating the ontological status of representations might help to support a constructivist perspective on cognitive theories.

References

- Churchland, P. (1989) *A Neurocomputational Perspective*. Cambridge, Massachusetts: MIT Press.
- Denett, D. (1991) *Consciousness Explained*. Boston: Little, Brown and Company.
- Edelman, G. M. (1992) *Bright Air, Brilliant Fire: On the Matter of Mind*. Basic Books.
- Fodor, J. A.; and Pylyshyn, Z. W. (1981) "How Direct is Visual Perception: Some Reflexions on Gibson's 'Ecological Approach'," *Cognition* 9.
- Gardner, H. (1985) *The Mind's New Science - A History of the Cognitive Revolution*. New York: Basic Books, Inc. Publishers.
- Gibson, J. J. (1966) *The Senses Considered as Perceptual Systems*. Boston: Houghton-Mifflin.
- Glaserfeld, E. von (1995) *Radical Constructivism*. Londond: The Falmer Press.
- Haugeland, J. (1985) *Artificial Intelligence: The Very Idea*. Cambridge, Massachusetts: MIT Press.
- Harnad, S. 1990. "The Symbol Grounding Problem", *Physica* D 42.
- Harnad, S. (1993) "The Origin of Words: A Psychophysical Hypothesis", W. Durham & B. Velichkovsky (Eds.), [Presented at ZIF conference

- on Biological and Cultural Aspects of Language Development], Muenster: Nodus Pub.
- Letvin, J. Y.; Maturana, H. R.; McCulloch, W. S.; and Pitts, W. H. (1988) "What the Frog's Eye Tells the Frog's Brain", *Embodiments of Mind*, Warren S. McCulloch. Cambridge, Massachusetts: MIT Press.
- Langer, S. (1948) *Philosophy in a New Key*. New York: Penguin Books.
- Llyod, Dan. (1989) *Simple Minds*. Cambridge, Massachusetts: MIT Press.
- McGee, V. (1991) *Truth, Vagueness, and Paradox*. Indianapolis: Hackett Publishing Company.
- Newell, A. 1990. *Unified Theories of Cognition*. MA: Harvard University Press.
- Putnam, H. (1981) *Reason, Truth, and History*. New York: Cambridge University Press.
- Quine, W. V. O. (1987) *Quiddities: An Intermittently Philosophical Dictionary*. Cambridge: Harvard University Press.
- Sander, J. (1996) "An Ecological Approach to Cognitive Science", *The Electronic Journal of Analytic Philosophy*, 4.
- Smith, B. (1996) *The Origin of Objects*, Cambridge, Massachusetts: MIT Press.
- Turvey, M.T.; Shaw, R. E. ; Reed, E. S. ; Mace, W. M. (1981) "Ecological Laws of Perceiving and Acting: In Reply to Fodor and Pylyshyn.," *Cognition* 9.