Representation in semiotics and in computer science

WINFRIED NÖTH

The concept of representation has been a key term of semiotics since the times of scholastic philosophy. In this long tradition, there have been many definitions and theories of representation, until, in our century, philosophers of postmodernism such as Foucault, Baudrillard, or Lyotard have come to the conclusion that we have reached a crisis of representation. Yet, at the same time when the philosophical idea of representation seemed thus to have fallen into a state of agony, computer science and artificial intelligence research have resumed the quest for the nature of representation in their project to model the representation of knowledge by means of intelligent machines. Hence, the topic of representation has even become a central issue of information science; but is the idea of representation, as pursued in computer science, the same as the one abandoned by postmodern philosophers? In fact, the term representation has been extremely polysemous both in the semiotic tradition and in computer science. Let us therefore begin with an investigation into the various meanings of representation in both domains of research.

Representation in the tradition of semiotic studies

In the tradition of semiotic studies, the concept of representation has been used to designate various aspects and types of signs and sign processes (cf. Nöth 1996). The most important ones are a sign in general, a sign vehicle, a mental concept or schema, a sign of the iconic category, and the sign relations of denotation or signification.

Representation as sign, a sign vehicle, or an interpretant

As a synonym of sign, the term representation appears, e.g., in the early writings of Peirce, who in 1865 distinguished between iconic, indexical,

Semiotica 115–3/4 (1997), 203–213 0037–1998/97/0115–0203

and symbolic 'representations' and defined semiotics as 'the general science of representations' (*W* 1.174, xxxii). A representation, alias sign, in this general sense is clearly distinguished from its three correlates described by Peirce as the representamen (the sign vehicle), the object (of reference), and the interpretant (or meaning) it creates in the interpreter's mind. However, the polysemy of the term representation has not stopped at this differentiation, for, in the narrower sense, the term representation is as commonly used to designate two of the correlates of the sign: in this narrower sense, a representation is either the sign vehicle, e.g., as a pictorial, dramatic, or verbal representation, or it is the interpretant in the interpreter's mind. In this latter sense, it has become a key term of cognitive science, which defines concepts or schemata as mental representations (cf. Nöth 1994). The roots of this mentalistic definition go back to the concept of *repraesentatio* in the semiotics of Enlightenment. Notice that Kant's '*Vorstellung*' is translated into English as 'representation'.

Representation as an iconic sign

The definition of representation as a sign of the iconic category predominated in medieval scholasticism. According to Duns Scotus, e.g., something that represents, *imitates* that which is represented by it (*repraesentativum imitatur suum repraesentatum*; cf. Kaczmarek 1986: 91). Today, the iconic concept of representation has been adopted by such different scholars as Nelson Goodman and Mario Bunge (cf. Nöth 1996).

Representation as a sign relation

The definition of representation as a sign relation appears in Peirce's later writings, when he began 'to confine the word representation to the operation of a sign or its relation to the object for the interpreter of the representation' (*CP* 1.540). 'To represent' is hence defined by Peirce as 'to stand for, that is, to be in such a relation to another that for certain purposes it is treated by some mind as if it were that other' (*CP* 2.273). Again we also find two more specific versions of this relational definition of representation. One is representation as the relation of denotation or the object relation of the sign. In this sense, Rosenberg (1981 [1974]: 1), e.g., defines designation as the verbal 'representation of things'. The other is the relation of signification or the interpretant relation, which we find, for example, in the view of representation as re-presentation, i.e., the process of bringing something previously present once more to the consciousness of a mind.

Representation in computer science

In which sense can computers be said to represent? Let us consider this question by investigating the use of the concept of representation in computer science.

Representation as sign, sign vehicle, or interpretant

The first and broadest possible definition is the one given by Stephen Palmer (1978: 262), who writes: 'A representation is, first and foremost, something that stands for something else. In other words, it is some sort of model of the thing (or things) it represents'. Here, representation is clearly a synonym of sign. The very wording of Palmer's definition is an echo of the medieval characterization of *signum as aliquid stat pro aliquo*. It is also in this tradition that the distinction between sign and sign vehicle is not clearly observed. If a representation is understood as that representing entity which represents something else, the term really defines a sign vehicle. This view of representation is also apparent in Allen Newell's paper on 'Physical symbol systems', which states that 'Representation is simply another term to refer to a structure that designates', where designation is specified as 'standing for' or 'aboutness', so that 'X *represents* Y if X designates aspects of Y, i.e., if there exist symbol processes that can take X as input and behave as if they had access to some aspects of Y' (Newell 1980: 176, 156).

In contrast to this view of representation as a sign vehicle of a sign, other computer scientists rather tend to conceive of representation as the interpretant of a sign. In this sense, Elaine Rich (1983: 136) describes the computer as the locus of internal representations, whose correlates, in the case of natural language processing, are on the one hand 'facts' and on the other hand sentences of a natural language. An internal computer representation of the fact expressed in the form of the English sentence 'Spot is a dog', e.g., would accordingly be its translation into the form 'Dog(Spot)' according to the rules of propositional logic. Irrespective of whether this logical form corresponds to any mental reality in the human mind, it is very much in accordance with Peirce's semiotics to consider such processes of translation from natural to artificial languages as processes of semiosis, and the resulting logical representation as an interpretant, since the interpretant, according to Peirce, is always a new sign which interprets or translates a previous sign. Between the two seemingly opposed views of representation, the one of the sign vehicle and the one of the interpretant, there is thus no real contradiction, but only a difference of focus on two necessarily related aspects, since the new sign generated as interpretant must by necessity also have a new sign vehicle which embodies this interpretant.

Representation as a symbol, icon, and index

Representation as symbolic sign. According to Newell's (1980) theory of the computer as a physical symbol system, computers represent knowledge by means of symbolic signs. Does Newell mean, by this thesis, symbols in the Peircean sense of an arbitrary and conventional sign? Often, Newell uses the term in a way which remains uncommitted as to the question of arbitrariness. When he writes that 'symbols are patterns that provide access to distal structures', Newell (1990: 77), uses the term merely as a synonym of sign, and the definition again echoes the medieval *aliquid pro aliquo* formula. Only in his paper with Simon does Newell explicitly raise the feature of arbitrariness to a criterion of symbolic representation, stating that 'given a symbol, it is not prescribed a priori what expression it can designate. This arbitrariness pertains only to symbols' (Newell and Simon 1981: 41).

Furthermore, Newell gives two other specifications which testify to a certain degree of arbitrariness in computer representation, namely, abstraction and codification. Abstraction means arbitrariness because of the incompleteness of the resulting representation. This feature is expressed in Newell's above-quoted condition that X representing Y means that X has access to [only] 'some aspects of Y' (see earlier section, 'Representation as a sign, a sign vehicle, or an interpretant). As a consequence of this only partial access which a representation can give of the knowledge it represents, Newell (1990: 80) concludes that 'symbol-level systems only approximate the knowledge level. Consequently, they do an imperfect job about the outside world'.

The second evidence of arbitrariness is apparent in what Newell (1990: 57–59) calls 'the law of representation'. It states that representation consists of encoding data from the external world into an internal system, where transformations are applied before the result is again decoded to be applied to an external situation. Of course, such processes of encoding and decoding involve the application of a code, which is, by definition, an arbitrary and conventional system correlating two sets of sign repertoires (cf. Nöth 1990a: 206–220). This view of representation as a kind of translation according to the rules of a code is expressed most clearly by Marr (1982: 20), who calls 'a representation ... a formal system for making explicit certain entities or types of information, together with a specification of how the system does this'. The example given in this context is the representation of numbers in the Arabic, Roman, and binary systems: the number thirty-seven, e.g., is represented in the Roman numeral system as XXXVII, in the binary numeral system as 100101.

Representation as a diagrammatic icon. Key words in the discussion of knowledge representation, such as mapping (Winograd and Flores 1986: 85), correspondence (Palmer 1978: 266–267), equivalence (cf. Jorna 1990: 33–35), or isomorphism suggest that iconicity is essential to the processes of representation. Such specifications of representation in terms of iconicity are not incompatible with, but complementary to, the ones who focus on representational symbolicity if we understand iconicity in the Peircean sense of diagrammatic iconicity which is based on the idea of structural correspondence or relational equivalence. The situation is analogous to the semiotics of language (cf. Nöth 1990b): the simplex morphological forms are essentially symbolic, while the grammatical patterns of sentences and texts are relational icons of the language structure. This relational aspect of iconicity in computer representation is most clearly elaborated by Palmer (1978: 266–267), according to whom 'the nature of representation is that there exists a correspondence (mapping) from objects in the represented world to objects in the representing world such that at least some relations in the represented world are structurally preserved in the representing world'. In terms of model theory, Palmer expresses this feature of diagrammatic iconicity as follows: 'A representational system is an ordered triple consisting of the two relational systems and a homomorphic function that maps the represented objects into the representing objects' (1978: 267).

The complementarity of symbolicity and iconicity in representation is even programmatic in Anderson's (1983: 45) tri-code theory of knowledge representation, which postulates 'three codes of representational types: a temporal string, which encodes the order of set items; a spatial image, which encodes spatial configuration; and an abstract proposition, which encodes meaning'. At a prototypical level, this model conceives of representation not only in terms of symbolicity, as in the case of abstract propositions, but also of iconicity (in the case of spatial images) and of indexicality, which is the sign type of temporal coding, since it encodes relations between points in time, which are always pointers both to past and future. Notice, however, that diagrammatic iconicity is involved in

Anderson's (1983: 47) description of all three types of encoding, since temporal string representation 'preserves temporal sequence' (which is a diagram), the spatial image 'preserves configurational information' and abstract propositions 'preserve semantic relations'.

Representation and indexicality. Indexical signs, which Peirce defined according to criteria such as causality and spatial or temporal contiguity, are the most ignored ones in the theory of computational representation. One of the problems is that such signs are often not considered as types of representation, and this is actually in accordance with the terminological tradition, where we only encounter iconic and symbolic definitions of representation.

Indexicality is the essence of computer semiosis at the level of physical causality and electronic connectivity. Electronic impulses generate patterns of electrical and magnetic activity, in which the impulse functions as a causal index of the pattern that it effects. Such processes of computer semiosis are often defined as nonrepresentational, in particular by theorists of connectionism and neural networks, yet Winograd and Flores do not hesitate to define these processes of indexical semiosis as the deepest level of representation in computers:

> Theoretically, one could describe the operation of a digital computer purely in terms of electrical impulses travelling through a complex network of electronic elements, without treating these impulses as symbols for anything. Just as a particular number in the computer might represent some relevant domain object (such as the location of a satellite), a deeper analysis shows that the number itself is not an object in the computer, but that some pattern of impulses or electrical states in turn *represents* the number. One of the properties unique to the digital computer is the possibility of constructing systems that cascade levels of representation on top of another to great depth. (Winograd and Flores 1986: 86–87)

At higher levels of computer semiosis, the index is again an essential sign type. The notions of command and immediate execution inherent in program expressions such as *assign, do, exit-if, continue-if* (Newell 1980: 144–145), the idea of goal directedness in terms like *immediate or indexed addressing* (Aho and Ullman 1992: 170), and the aspect of causal connectivity between an *input* and its *output* testify to the dimension of indexicality at the programming level.

The aspect of causal dependency between the two structures is a feature according to which there is an element of indexicality in any relation between a representing and a represented domain in computer semiosis. The idea appears as a criterion in Newell's (1980: 156) definition of designation: 'An entity X designates an entity Y relative to a process P, if, when P takes X as input, its behavior depends on Y'. The aspect of indexicality in such dyadic relations of dependency is even more apparent when Newell (1990: 74–75) begins to interpret the relations between the symbol structures X and Y in localistic terms. X and Y are structures localized in different regions of the physical symbol machine: 'The symbol token [X] is the device in the medium that determines to go outside the local region to obtain more structure'. It provides 'access to the distal structure' Y which is 'transported by *retrieval* from the distal location to the local site' (1990: 74–75). All these descriptions are characterizations of indexical modes of semiosis.

The dyadic sign model: Signification

The predominant sign model in the theory of computational representation is a dyadic one. For example, Palmer (1978: 262), in his definition of representation, postulates 'the existence of two related but functionally separated worlds: the *represented world* and the *representing world*', and states that 'the job of the representing world is to reflect some aspects of the represented world in some fashion'. In Newell's model of symbolic representation, consisting of the dyad X and Y, X is a symbol which designates (or gives access to) the symbol Y. Such dyads of signs in which one of the two gives information about, or interprets, the other, are clearly instances of relations between sign (vehicle) and interpretant, in other words, relations of signification.

The sign and its object: Denotation

The question remains to be raised whether computer signs, like language signs, have any relation to objects of the world for which they might stand in some way, in other words, whether computer semiosis has a dimension of denotation besides the one of signification.

The question of the object relation of the sign has been quite controversial in the history of semiotics. In the tradition from the Epicureans via Port Royal and Saussure to the Radical Constructivists of today, the sign has been dogmatically dyadic, excluding by definition the concepts of denotation or reference, whereas the sign had been triadic in a tradition extending from the Stoics to Peirce.

Without being able to go too far into detail, we will adopt, in the following, the Peircean view, according to which the object can be both an 'object of the world' with which we have a 'perceptual acquaintance'

(*CP* 2.330) and a merely mental or imaginary entity 'of the nature of a sign or thought' (*CP* 1.538). The object, in this view, is some knowledge of the world which we have before it is instantiated in a given instance of semiosis when it gives rise to a new interpretation called interpretant. While the interpretant is thus directed toward the future of the sign in the process of semiosis, the object is the sign's representational dimension of the past (cf. Santaella 1995: 59). With Peirce, it is furthermore necessary to distinguish between two kinds of objects, the immediate and the dynamical object. The immediate object corresponds to the interpreter's previous knowledge of, or acquaintance with, the object, no matter whether such an object exists 'in reality' or not (*CP* 4.536). The dynamical object, by contrast, is the 'object outside of the sign' (Hardwick 1977: 83) in the sense of a 'reality which by some means contrives to determine the sign to its representation' (*CP* 4.536), or that which the sign 'can only indicate and leave the interpreter to find out by collateral experience' (*CP* 8.314). In other words, while the immediate object is a representation of the object in a mind, the dynamical object describes a pragmatic or experiential acquaintance with the object in the mind's interaction with it.

Against this background, the question of the object relation in computer semiosis excludes such naive views of representation as those which substitute the sign-interpretant dyad by a naive sign-object dyad. Some computer scientists, in fact, have defined the concept of representation in this latter sense. For Bobrow (1975: 2–3), for example, representation is 'the result of a selective mapping of aspects of the world', which correlates a 'world-state' with a 'knowledge-state', and Haugeland (1981: 22), in his discussion of semantic engines, speaks of signs as 'relations which connect ... tokens to the outside world'.

Such accounts of the sign in computer semiosis, which so to speak endow the computer with a 'window to the world', are not only naive from the point of view of semiotics, but have also been attacked by other computer scientists, who, like Winograd and Flores (1986: 85) argue that it is not the computers but only the 'community of programmers, who can know what kind of external object a computer symbol may represent: 'The problem is', they argue, 'that representation is in the mind of the beholder. There is nothing in the design of the machine or the operation of the program that depends in any way on the fact that the symbol structures are viewed as representing anything at all' (1986: 86).

I have already shown earlier that we have to disagree with the thesis that computers do not represent, if representation is taken in the sense of signification; but do we also have to disagree if representation is taken in the sense of denotation? The answer depends on the kind of task a computer is able to perform. As long as the task is restricted to mere semantic, propositional, or even pictorial representation, we must indeed conclude that such processes basically only involve signification and no denotation. They are essentially processes of transcoding without any window to the world. Since they consist of processes of signs transformed again into signs, one might be tempted to interpret these processes as self-referential.

Yet, how about computers who interact not only with programmers and users, but also with their physical environment such as robot-like machines with sensors and effectors operating in the physical world (cf. Winograd and Flores 1986: 86)? The standard dogma of computer philosophy defended, among others, by Searle (1981) and Fodor (e.g., 1986: 11), is that robots, and, by the way, also paramecia, cannot have mental representations, neither in the significational, nor in the denotational sense because they are nonintentional systems. However, intentionality, defined as the causal property of having beliefs, desires, thoughts, and so forth, is an extremely high threshold for the description of semiotic processes, and it is even questionable whether human semiosis can be reduced to such states. In a semiotics like the one of Peirce, which, in contrast to intentional semantics, does not share the premise of intentionality as a criterion of representation and semiosis, the question whether robots can have a denotational representation of their environment is once more open. Let us consider it in the light of the Peircean definition of the object.

The robot we are thinking of is an autonomous mobile agent designed to move around in space and thereby to avoid obstacles in its environment. Our robot has perceptual modules which give a symbolic representation of the world and also action modules which generate and execute the desired movement in its environment (cf. Brooks 1991: 146–147). Let us further assume that our robot is not only able to sense the objects in its environment according to its program, that it is able to avoid hitting them by halting and moving away from them, but also to learn from previous experience by changing the symbolic representation of its environment in its perception module, to cope more successfully with its environment in the future.

I would like to argue that this computer is provided with the capability for representation in the sense of denotation. The robot starts out with a program in its perception module which represents a world of immediate objects. It is true that the perceptual acquaintance of the environment with which this robot is equipped is not the result of its own previous experience, but of the programmer's simulation of such an experience. However, as soon as the robot interacts with, and learns from, its actual environment, it enters into contact with a world of dynamical objects.

The actual reality of these objects is not yet represented as such within the perceptual module but can become represented in it in a process of trial and error. The interpretative result computed by the robot in its interactions with the dynamical objects of its environment is the interpretant in this process of semiosis. It is able to affect the robot's future interpretations of its environment. In future situations, the thus acquired knowledge is available as the immediate object in the robot's interaction with its environment.

The crisis of representation?

I have started out with a brief remark on the crisis of the idea of representation in postmodern philosophy. This topic, in fact, covers a wide range of different critical approaches such as Lucács's thesis of the impossibility of representation in the art of the twentieth century, Foucault's theory of the loss of representation in philosophical discourse since the nineteenth century, or Derrida's critique of the idea of representation as a 're-presentation'. The various lines of argument cannot be dealt with in depth here (but see Nöth 1996). In the context of our topic, however, it is of interest to note that a crisis of representation seems also to have reached the theoretical debate in computer science. Not only do connectionists and advocates of parallel processing postulate the possibility of intelligent machines without representation (cf. Brooks 1991), but also the concept of representation has been in some crisis with those who have used it without a solid semiotic foundation. I hope to have shown that the theory of representation requires such a semiotic foundation and that semiotics may thus contribute to overcome the crisis that has begun to affect the discourse of computer science.

References

Aho, Alfred V. and Ullman, Jeffrey D. (1992). *Foundations of Computer Science*. New York: Freeman.

Anderson, John R. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.

Bobrow, Daniel G. (1975). Dimensions of representation. In *Representation and Understanding*, D. G. Bobrow and A. Collins (eds.), 1–34. New York: Academic Press.

Brooks, Rodney A. (1991). Intelligence without representation. *Artificial Intelligence* 47, 139–159.

Fodor, Jerry A. (1986). Why paramecia don't have mental representations. *Midwest Studies in Philosophy* 10, 3–23.

Hardwick, Charles S. (ed.) (1977). *Semiotic and Significs: The Correspondence between Charles S. Peirce and Victoria Lady Welby*. Bloomington: Indiana University Press.
Haugeland, John (1981). Semantic engines. In *Mind Design*, J. Haugeland (ed.), 1–34. Cambridge, MA: MIT Press.
Jorna, René (1990). *Knowledge Representation and Symbols in the Mind*. Tübingen: Stauffenburg.
Kaczmarek, Ludger (1986). Quid sit aliquid repraesentari in verbo. In *Geschichte und Geschichtsschreibung der Semiotik*, Klaus D. Dutz and Peter Schmitter (eds.), 85–110. Münster: MAKS.
Marr, David (1982). *Vision*. New York: Freeman.
Newell, Allen (1980). Physical symbol systems. *Cognitive Science* 4, 135–183.
—(1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.
Newell, Allen and Simon, Herbert A. (1981). Computer science as empirical enquiry. In *Mind Design*, John Haugeland (ed.), 35–66. Cambridge, MA: MIT Press.
Nöth, Winfried (1990a). *Handbook of Semiotics*. Bloomington: Indiana University Press.
—(1990b). The semiotic potential for iconicity in spoken and written language. *Kodikas/Code* 13, 191–209.
—(1994). Semiotic foundations of the cognitive paradigm. *Semiosis*, 73, 5–16.
—(1996). Signo, representação e representação mental. In *Encontro com as Ciencias Cognítivas*, María Euníce Gonzales et al. (eds.), 53–85. São Paulo: UNESP.
Palmer, Stephen E. (1978). Fundamental aspects of cognitive representation. In *Cognition and Categorization*, Eleanore Rosch and Barbara B. Lloyd (eds.), 259–303. Hillsdale, NJ: Erlbaum.
Peirce, Charles S. (1931–1958). *Collected Papers of Charles Sanders Peirce*, Charles Hartshorne, Paul Weiss, and A. W. Burks (eds.), vols. 1–8. Cambridge, MA: Harvard University Press. [Reference to Peirce's papers will be designated *CP*].
—(1982–). *Writings of Charles S. Peirce: A Chronological Edition*, M. Fisch, E. Moore, and C. Kloesel (eds.), vols. 1–5. Bloomington: Indiana University Press. [Reference to Peirce's writings will be designated *W*.]
Rich, Elaine (1983). *Artificial Intelligence*. New York: McGraw-Hill.
Rosenberg, Jay F. (1981 [1974]). *Linguistic Representation*. Dordrecht: Reidel.
Santaella [Braga], Lucia (1995). *A teoria geral dos signos*. São Paulo: Ática.
Searle, John R. (1981). Mind, brains, and programs. In *Mind Design*, John Haugeland (ed.), 282–306. Cambridge, MA: MIT Press.
Winograd, Terry and Flores, Fernando (1986). *Understanding Computers and Cognition*. Norwood, NJ: Ablex.

Winfried Nöth (b. 1944) is Professor of English Linguistics and Semiotics at the Universität Gesamthochschule Kassel (Germany), Dean of the Department of Modern Languages, and permanent Visiting Professor of Semiotics at São Paulo Catholic University ⟨noeth@hrz.uni-kassel.de⟩. His principal research interests are in semiotics and semiotic linguistics. Among his publications are *Strukturen des Happenings* (1972), *Dynamik semiotischer Systeme* (1977), *Literatursemiotische Analysen zu Lewis Carroll's Alice-Büchern* (1980), *Handbook of Semiotics* (1990), *Origins of Semiosis* (1994), *Panorama da Semiótica de Platão a Peirce* (1995), *A semiótica no século XX* (1996), and *Semiotics of the Media* (forthcoming 1997).