

SYMBOLIC, CONCEPTUAL AND SUBCONCEPTUAL REPRESENTATIONS

Peter Gärdenfors

Cognitive Science, Lund University
Kungshuset, Lundagård
S-223 50 Lund, Sweden

THE PROBLEM OF REPRESENTING INFORMATION

Cognitive science aims at understanding how information is represented and processed in different kinds of agents, biological as well as artificial. The research has two overarching goals. One is *explanatory*: By studying the cognitive activities of humans and other animals, one formulates *theories* of different kinds of cognition. The theories are tested either by experiments or by computer simulations. The other goal is *constructive*: By building *artifacts* like chess-playing programs, robots, animats, etc, one attempts to construct systems that can solve various cognitive tasks. For both kinds of goals, a key problem is *how the information* used by the cognitive system *is to be modelled* in an appropriate way.

Within cognitive science, there are currently two dominating approaches to the problem of representing information. (I prefer to talk about representing *information* rather than representing knowledge, since this term is much more natural for the conceptual and subconceptual forms of representation. Furthermore, the term is much less philosophically loaded). The *symbolic* approach starts from the assumption that cognitive systems should be modelled by Turing machines. The second approach is *connectionism*, which models cognitive systems by artificial neuron networks. Both of these approaches have their advantages and disadvantages. They are often presented as competing paradigms, but since they attack cognitive problems on different levels, they should rather be seen as complementary methodologies.

However, as I shall argue, there are aspects of cognitive phenomena for which neither symbolism nor connectionism seem to offer appropriate modelling tools. In this article, I will advocate a third form of representing information that is based on using *geometric* structures rather than symbols or connections between neurons. I shall call it *conceptual* representations since I believe that the essential aspects of concept formation should be described on this level. Again, conceptual representations should not be seen as competing with symbolic or connectionistic representations. Rather, the three kinds can be seen as three *levels* of

representations of cognition. Or, since all levels may be present in one system, they may as well be called *perspectives*.

The main thesis in this paper is that all three levels are needed in order to cover the problems of representation one faces when explaining cognitive phenomena and when building artificial agents. In particular, I submit that the conceptual level is necessary to explain how symbolic representations can arise from connectionist, or, more generally, *subconceptual*, representations.

After presenting and criticising the symbolic and connectionists representations, I shall outline a theory of *conceptual spaces* as a particular framework for representing information on the conceptual level. The theory of conceptual spaces should not primarily be seen as an empirical theory, but as a tool for constructing representations in artificial systems. However, I believe that it also makes some sense in relation to what is known about representations in biological systems. Finally, I shall argue that the three different forms of representations are connected with different computational methodologies.

SYMBOLIC REPRESENTATIONS

Computationalism

The outline of the symbolic paradigm of representing information to be presented here will not be explicitly found in the works of any particular author. It forms an implicit methodology for most research in AI. The classical sources are the works of Newell and Simon.¹ More recently, a defense of the general reasoning can be found, for example, in the writings of Jerry Fodor and Zenon Pylyshyn.^{2,3,4} Since the position is well-known, a sketch of the most relevant features will suffice for my purposes.

According to the paradigm, the atoms of representations are *symbols* which combine to form meaningful expressions. Here I am referring to the traditional sequential kind of computer programs with “explicit” symbol representations and not to parallel distributed processing which may use “intrinsic” representations.⁵ Such systems and their representations will be discussed below.

Within the symbolic tradition there are two main application areas that go hand in hand: one is modelling logical *inferences* and the other is syntactical *parsing*. When the symbols are used for modelling logical inference, the expressions represent *propositions* and they stand in various *logical relations* to each other. Information processing involves above all *computations* of logical consequences. In brief, a cognitive agent is seen as a kind of logic machine that operates on sentences from some formal language.

The central tenet of the symbolic paradigm is that representing and processing information essentially consists of *symbol manipulation*.⁶ The manipulations of symbols are performed without regards to the semantic content of the symbols.

The symbols can be concatenated to form expressions in a *language of thought* – sometimes called Mentalese. The content of a sentence in Mentalese is a belief or a thought of an agent. The different sentential or propositional attitudes in the cognitive states of a person are connected via their *logical or inferential relations*. Pylyshyn writes: “If a person believes (wants, fears) P, then that person’s behavior depends on the form the expression of P takes rather than the state of affairs P refers to ...”.⁷ In applications within AI, first order logic has been the dominating inferential system, but in other areas more general forms of inference, like those provided by inductive logic or decision theory, have been utilized.

Processing the information contained in a cognitive state consists in *computing* the consequences of the sentential attitudes, using some set of inference rules. The following quotation from Fodor is a typical formulation of the symbolic paradigm:

“Insofar as we think of mental processes as computational (hence as formal operations defined on representations), it will be natural to take the mind to be, inter alia, a kind of computer. That is, we will think of the mind as carrying out whatever symbol manipulations are constitutive of the hypothesized computational processes. To a first approximation, we may thus construe mental operations as pretty directly analogous to those of a Turing machine.”⁸

Similarly, the Chomsky tradition in linguistics focuses on the *syntax* of language. Language is seen as strings of formal symbols that can be processed by different kinds of *automata*, of which the Turing machine is the most advanced. The main operations are *parsing* of a string of symbols according to a (recursive) set of recursive grammatical rules, and, conversely, *generation* of strings according to the grammatical rules.

The *material basis* for the symbolic processes, be they logical, linguistic or of a more general psychological nature, is irrelevant to the description of their results. The inference rules of logic and the electronic devices which conform to these rules are seen to be analogous to the workings of the brain. In brief, the mind is thought to be a computing device, which generates symbolic sentences as inputs from sensory channels, performs logical operations on these sentences, and then transforms them into linguistic or non-linguistic behaviors as output.

The limitations of symbolic representations

After this outline of the position, I will now turn to the limitations of the representational power of the symbolic approach. One of the major problems encountered in the classical form of AI is the *frame problem*.^{9,10} It was hoped that if we could formulate the knowledge necessary to describe the world and the possible actions in a suitable symbolic formalism, then by coupling this world description with a powerful inference machine one could construct an artificial agent capable of planning and problem solving. The frame problem can be defined as the problem of specifying in symbolic formalism what changes and what stays constant in the particular domain where the agent is acting.

It soon turned out, however, that describing actions and their consequences led to a combinatorial explosion. The main problems are connected with describing what in the world remains unaffected when a particular action is performed. Some changes are *relevant* for the planning, others totally irrelevant. Propositional representations are not well suited for representing *causal connections*. One of the main reasons for this is that it does not provide any natural way to separate different *domains* of information (this point will be elaborated below).

The frame problem is connected to the fact that the central bearers of the symbolic representations based on first order languages are the *predicates* of the language. These predicates are supposed to be *given* to the system. Stewart says: “In the computational paradigm, symbolic representations are theoretical primitives so that it is not really possible to study their evolutionary emergence, because there are no conceptual categories available for specifying the situation *before* symbols came into being.”¹¹ However, a successful system must be able to learn radically new properties from its interactions with the world, and not only form new combinations of the given predicates. So a crucial question for a theory of cognitive representation is: *where do new predicates come from?*

Furthermore, not only is there a problem of describing the genesis of predicates, but their *development* in a cognitive system is not easily modelled on the symbolic level. Even after an agent has learned a concept, the meaning of the concept very often *changes* as a result of new experiences. In the symbolic mode of representation, there has been no successful way of modelling the dynamics of concepts. The fact that artificial neuron networks can adapt their categorizations to new experiences has been claimed as an advantage of the networks over symbolic systems, but I believe that the conceptual level is the right one to handle this kind of process.

Most adherents of the symbolic paradigm are semantic *realists* in the sense that the “meaning” of a predicate or a sentence is determined by mapping it to the external world (or, to make it even more remote from a cognitive system, to a plethora of possible worlds). The world (and the mapping) is assumed to exist independently of any relation to a cognitive subject. A clear example of this position is given by Fodor: “If mental processes are formal [symbolic], then they have access only to the formal properties of such representations of the environment as the senses provide. Hence, they have no access to the semantic properties of such representations, including the property of being true, of having referents, or, indeed, the property of being representations of the environment”¹² and “We must now face what has always been the problem for representational theories to solve: what relates internal representations to the world? What is it for a system of internal representations to be semantically interpreted?”¹³ These problems arise for the symbolic paradigm because it operates with a realist semantics that presume external representations.

This view on the semantics of the symbols makes it difficult to explain how the meanings of the predicates *change* during the cognitive development of an agent. Semantic realists are more or less obliged to assume that the meanings of symbols are fixed. In a sense, this view of semantics is inherited from the model theory of mathematical logic. For mathematical concepts, however, we never have the problem of adapting concepts to new encounters with reality.

Another problem for the symbolic approach is highlighted by Harnad.¹⁴ He asks the following questions: “How can the semantic interpretation of a formal symbol system be made *intrinsic* to the system, rather than just parasitic on the meanings in our heads? How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols?” This problem he calls the *symbol grounding problem*. He says: “But the problem of connecting up with the world in the right way is virtually coextensive with the problem of cognition itself.” The symbol grounding problem can be argued to be an *artifact* of the symbolic position. In the same vein, Stewart says:

“[...] since linguistic symbols emerge from the precursors of the semiotic signals of animal communication, they *always already* have meaning, even before they acquire the status of symbols. On this view, formal symbols devoid of meaning are *derivative*, being obtained by positively *divesting* previously meaningful symbols of their significance. Quite concretely, this process occurred historically in the course of the history of axiomatic mathematics from Euclid to Hilbert. From this point of view, the “symbol-grounding problem” of computation cognitive science looks rather bizarre and somewhat perverse: why go to all the bother of divesting “natural symbols” of their meaning, and then desperately trying to put it back, when it would seem so simple to leave them as they are!”¹⁵

These problems have been swept under the carpet within the symbolic tradition. And among those who have addressed the problem no satisfactory solution has been provided.

The problems concerning the formation and dynamics of predicates become most pressing when one scrutinizes the attempts within the symbolic tradition to explain *inductive inferences*. The most ambitious project of analyzing induction during this century has been that of the logical positivists. Inductive inferences were important for them, since such inferences were necessary for their verificationist aims. The basic objects of study for them were sentences in some more or less regimented language. Ideally, the language was a version of first order logic where the atomic predicates represented observational properties. These observational predicates were taken as *primitive*, unanalysable notions. The main tool used when studying the symbolic expressions was logical analysis.

However, it became apparent that the methodology of the positivists led to serious problems in relation to the problem of induction. The most famous ones are Hempel’s¹⁶ “paradox of confirmation” and Goodman’s¹⁷ “riddle of induction.” What I see as the root of the troublesome cases is that if we use *logical* relations alone to determine which inductions

are valid, the fact that all predicates are treated on a par induces *symmetries* which are not preserved by our understanding of the inductions: “Raven” is treated on a par with “non-raven,” “green” with “grue” etc. What we need is a *non-logical* way of distinguishing those predicates that may be used in inductive inferences from those that may not.

There are several suggestions for such a distinction in the literature. One idea is that some predicates denote “natural kinds” or “natural properties” while others don’t, and it is only the former that may be used in inductive reasoning. Natural kinds are normally interpreted realistically, following the Aristotelian tradition, and thus assumed to represent something that exists in reality independently of human cognition. However, when it comes to inductive *inferences* it is not sufficient that the properties exist out there somewhere, but we need to be able to grasp the natural kinds by our minds. In other words, what is needed to understand induction, as performed by humans, is a *conceptualistic* analysis of natural properties.

Even though AI researchers have had some success in their attempts to mechanize induction, it is clear that their methodology suffers from the same general problems as the symbolic level in general. The enigmas of induction that have been unearthed by Goodman, Hempel and others are also applicable to the induction programs in recent mainstream AI.

Trying to capture inductive inferences by an algorithm also highlights some of the general limitations of the symbolic representations. The programs work by considering the applicability of various logical combinations of the atomic predicates. But the epistemological origin of these predicates is never discussed. Even though AI researchers are not actively defending the positivist methodology, they are following it implicitly by treating certain predicates as observationally, or at least externally, given. However, the fact that the atomic predicates are assumed as granted from the beginning means that much inductive processing has already been performed.

The core of the problem, it seems to me, is that the symbolic level is insufficient for handling the problems of induction and similarity. We not only want to know how observational predicates should be combined in the light of inductive evidence, but, much more importantly, *how the basic predicates are established* in the first place. This problem has, more or less, been neglected by the logical positivists. Logical analysis, the prime tool of positivism, is of no avail for these forms of concept formation. In brief, the symbolic approach to induction sustains no creative inductions, no genuinely new knowledge, and no conceptual discoveries. To do this, we have to go below the symbolic level.

In my opinion, the symbolic paradigm has a very limited applicability. For most areas of representation and information processing, including semantic representation, it is positively misleading. I do not claim that the symbolic paradigm is totally without value: if one is trying to imitate natural language understanding, it is necessary to use some linguistic structures, for example, in order to be able to analyse the grammar of the input. But even in this symbol-oriented area we encounter problems when it comes to providing a *semantics* for the linguistic expressions. According to the symbolic approach, the content of an expression in a natural language would be represented by an expression in Mentalese. But this would basically be a *translation* from one language to another and it would not help us understand how the expression gets its *meaning*.

THE SUBCONCEPTUAL LEVEL

Connectionism

Connectionist systems, also called artificial neuron networks (ANNs), consist of large numbers of simple but highly interconnected units (“neurons”). The units process information in *parallel*, in contrast to most symbolic models where the processing is serial.

(The distinction between serial and parallel processing is not, in itself, crucial for the representational powers of a model.) There is no central control unit for the network, but all neurons “act” as individual processors. Hence connectionist systems are examples of *parallel distributed processes* (PDP).¹⁸

Each unit receives activity, both excitatory and inhibitory, as input; and transmits activity to other units according to some function (normally non-linear) of the inputs. The behavior of the network as a whole is determined by the initial state of activation and the connections between the units. The inputs to the network also change the “weights” of the connections between units according to some learning rule. Typically, the change of connections is much slower than changes in activity values. The units have no memory of themselves, but earlier inputs may be represented indirectly via the changes in weights they have caused. In the literature one finds several different kinds of connectionist models that can be classified according to their architecture or their learning rules.

According to connectionism, cognitive processes should not be represented by symbol manipulation, but by the *dynamics* of the patterns of activities in ANNs.

Representations in connectionist and related systems

Palmer¹⁹ introduces a distinction between *intrinsic* and *extrinsic* representation. Representation is intrinsic when the representing relation has the same inherent constraints as its represented relation. For example, if the age of a class of objects is represented by the height of rectangles, the structure of the represented relation (age) is intrinsic in the representing relation (height). In contrast, representing age by numbers is an extrinsic representation since the structure of the digit sequences does not have the same structure as the represented relation. Intrinsic representations *resemble* what they represent. In contrast, extrinsic representations must be accompanied by a *rule* which specifies how the representation is to be interpreted – such a rule provides the “meaning” of the representation. Representations in connectionist systems are generally intrinsic, while in symbolic models they are, almost by definition, extrinsic.

Connectionist systems have become popular among psychologists and cognitive scientists since they seem to be excellent tools for building models of associationist theories. And networks have been developed for many different kinds of tasks, including vision, language processing, concept formation, inference, and motor control. Among the applications, one finds several that traditionally were thought to be typical symbol processing tasks. In favor of the neural networks, it is claimed by the connectionists that these models do not suffer from the brittleness of the symbolic models and that they are much less sensitive to noise in the input.

Given that we are focussing on the representational aspects of cognitive systems, let us then consider the information on the subconceptual level. How do we distill sensible information from what is received by a set of receptors? Or, in other words, how do we make the transition from the subconceptual to the conceptual and the symbolic levels? These questions point to the representation problems that occur on the subconceptual level.

The basic problem is that the information received by the receptors is too rich and unstructured. What is needed is some way of transforming and organizing the input into a form that can be handled on the conceptual or linguistic level. There are several methods for treating this kind of problem. Within the area of ANNs, there are systems which are developed to perform this kind of dimensionality reduction, e.g., Kohonen’s²⁰ self-organising networks.

What are then the drawbacks of using neural networks of the type described here for information representation? A fundamental epistemological problem is that even if we know that the network has learned to categorize the input in the right way, we may not be able to describe *what* the emerging network represents. This kind of level problem is ubiquitous in

applications of neural networks for learning purposes. The upshot is that a future theory of neural networks must somehow bridge the gap of going from the subconceptual level to the conceptual level. We may account for the information provided at the subconceptual level in term of a dimensional space with some topological structure, but there is no general recipe for determining the *conceptual* meaning of the dimensions of the space.

THE CONCEPTUAL LEVEL

Conceptual spaces

A crucial question for any theory of representation is how concepts should be modelled. On the symbolic level, basic concepts are not modelled, just *named* by the basic symbols. Then names of more complex concepts are constructed by compositions, logical or syntactical, of the simple names. And when it comes to connectionist systems, concepts are often represented *implicitly* in such systems, as was argued in the previous section. The primary motivation for introducing a conceptual level is to provide tools for *explicit* representations of basic concepts. These representations will be taken to be the *references* of the symbols. It is natural to view the conceptual level as being between the symbolic and the subconceptual levels.

When attacking the problem of representing concepts, an important aspect is that the concepts are not independent of each other but can be structured into *domains*: Spatial concepts belong to one domain, concepts for colors to a different domain, kinship relations to a third, concepts for sounds to a fourth, and so on.

The central claim of this paper is that the notion of a *conceptual space* is a fruitful way of modelling such domains. A conceptual space consists of a number of *quality dimensions*. As examples of quality dimensions one can mention temperature, weight, brightness, pitch and the three ordinary spatial dimensions height, width and depth. I have chosen these examples because they are closely connected to what is produced by our sensory receptors.²¹ The spatial dimensions height, width and depth as well as brightness are perceived, by the visual sensory system, pitch by the auditory system, temperature by thermal sensors and weight, finally, by the kinesthetic sensors. However, there are also quality dimensions that are of an abstract non-sensory character. The primary function of the quality dimensions is to represent various “qualities” of objects. They form the “framework” used to assign *properties* to objects and to specify *relations* between them.

The dimensions are taken to be independent of symbolic representations in the sense that we and other animals can represent the qualities of objects, for example when planning an action, without presuming an internal language or another symbolic system in which these qualities are expressed. In other words, the dimensions are the building blocks of representations on the conceptual level. The quality dimensions should be seen as abstract representations used as a modelling factor in describing mental activities of organisms, and sometimes also activities of artificial systems. They are thus not required to have any immediate physical realisation.

The notion of a dimension should be understood literally. It is assumed that each of the quality dimensions is endowed with a certain structure. For most of the examples of quality dimensions, these structures will be of a *geometric* nature (in some cases they are *topological* or *orderings*). As a first example to illustrate such a structure, I will take the dimension of “time”. In science, time is modelled as a one-dimensional structure which is isomorphic to the line of real numbers. If “now” is seen as the zero point on the line, the future corresponds to the infinite positive real line and the past to the infinite negative line.

This representation of time is not universal, but is to some extent culturally dependent, so that other cultures have a different time dimension as a part of their cognitive structure.

For example, in some cultural contexts, time is viewed as a *circular* structure. There is thus no unique way of choosing a dimension to represent a particular quality, but in general a wide array of possibilities.

Another example is the dimension of “weight” which is one-dimensional with a zero point, and thus isomorphic to the half-line of non-negative numbers. A basic conceptual constraint on this dimension is that there are no negative weights.

It should be noted that some quality “dimensions” have only a *discrete* structure, that is, they merely divide objects into disjoint classes. Two examples are classifications of biological species and kinship relations in a human society. However, even for such dimensions one can distinguish a simple geometric structure. For example, in the phylogenetic classification of animals, it is meaningful to say that birds and reptiles are more *closely related* than reptiles and crocodiles.

In previous writings on conceptual spaces,^{22,23,24} I have used the example of the perceptual color space to illustrate a more structured set of quality dimensions. However, one can also find related spatial structures for other sensory qualities. For example, consider the quality dimension of pitch, which is basically a continuous one-dimensional structure going from low tones to high. This representation is directly connected to the neurophysiology of pitch perception. The cochlea of the inner ear functions so that high frequency tones stimulate receptor cells at the base of the cochlea, and lower tones stimulate cells higher up in the spiral. In this way the positions in the cochlea map, in a logarithmic fashion, the frequencies of the sounds received by the ear. Thus acoustic frequency is spatially coded in the nervous system. This is a paradigm example of an intrinsic representation in the sense of Palmer.

Apart from the basic frequency dimension of tones, we can find some interesting further structure in the mental representation of tones. Natural tones are not simple sinusoidal tones of only one frequency, but constituted of a number of higher harmonics. The timbre of a tone is determined by the relative strength of the higher harmonics of the fundamental frequency of the tone. An interesting perceptual phenomenon is “the case of the missing fundamental.” If the fundamental frequency is removed by artificial methods from a complex tone, the pitch of the tone is still perceived as that corresponding to the removed fundamental.²⁵ Apparently, the fundamental frequency is not indispensable for pitch perception, but the perceived pitch is determined by a combination of the lower harmonics.

Thus, the harmonics of a tone are essential for how it is perceived. This entails that tones which share a number of harmonics will be perceived to be similar. The tone that shares the most harmonics with a given tone is its octave, the second most similar is the fifth, the third most similar is the fourth and so on. This additional “geometric” structure on the pitch dimension, which can be derived from the wave structure of tones, provides the foundational explanation for the perception of musical *intervals*.²⁶

For another example of sensory space representations let me mention that the human perception of *taste* appears to be generated from four distinct types of receptors: salt, sour, sweet, and bitter. Thus the quality space representing tastes could be described as a 4-dimensional space. One such model was proposed by Henning in 1916,²⁷ who suggested that gustatory space could be described as a tetrahedron as in figure 1.

For most kinds of dimensions it will be possible to talk about *distances*. The *similarity* of two objects can therefore be defined via the distance between their representing points in the space. Thus conceptual spaces provide us with a natural way of representing similarities, which has turned out to be very difficult to handle on the symbolic level.

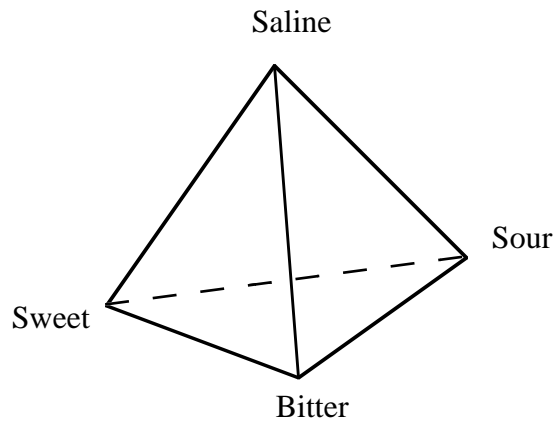


Figure 1. Henning's proposal for the structure of gustatory space.

A *property* can then be represented as a *region* of a conceptual space. The use of regions heavily exploits the geometric or topological structure of the spaces. Furthermore an *object* can be represented as a point in a conceptual space. Such a point may belong to several of the regions that represent properties. In this way the object is directly represented as having a number of properties. A *concept* can be defined as a just a property, i.e., a region of a space or as a set of properties from different dimensions.

One of the main applications of the theory is to use conceptual spaces for representing the *semantics* of symbolic expression. The meanings of different kinds of expressions will be defined in terms of different constructions of elements from the spaces. Since these constructions will be dependent on the structure of the underlying quality dimensions, it follows that the assignment of meanings to the expressions on the symbolic level is far from arbitrary. On the contrary, the semantics (and to some extent even the grammar) of the linguistic constituents will be severely constrained by the underlying structure. This is anathema for the Chomskian tradition within linguistics, but, as a matter of fact, it is one of the central tenets of the recently developed *cognitive linguistics*.^{28,29}

The advantages of the conceptual level

What are the advantages of focussing on the conceptual level and using conceptual spaces as a tool for representing information? With regard to the problems for symbolic representations that were presented earlier it should first be noted that conceptual spaces will solve the symbol grounding problem. The symbols are given meaning by being connected to various constructions in the spaces. The resulting semantics are of a cognitive kind since, unlike a realist semantics, it does not presume any objects outside the cognitive structure to determine the meaning of the expressions of language. The external world only enters the picture when the truth or validity of the expressions are to be evaluated.

Conceptual spaces can also provide a better way of representing learning in general and concept formation in particular than what can be achieved on the symbolic level. Many of the problems of induction that are created by the symbolic approach dissolve into thin air when analysed on the conceptual level.³⁰ Similarly, the problem of how transducers work becomes a non-problem since no transducers are needed for the information represented in conceptual spaces.

The theory of conceptual spaces may also indicate a direction where a solution to the frame problem can be ferreted out. The starting point is to separate the information to be represented into *domains*. The combinatorial explosion of symbolic representations of a changing world is a result of not keeping symbolic information about different domains separated.

One notion that is severely downplayed by the symbolic representations is that of *similarity*. However, I submit that judgements of similarity are central for a large number of cognitive processes. Such judgements reveal the dimensions of our perceptions and their structures. Quine discusses similarity and its relation to that of a natural kind.³¹ He notes that it “is immediately definable in terms of kind; for things are similar when they are two of a kind.” Furthermore, he says about similarity that we

“cannot easily imagine a more familiar or fundamental notion than this, or a notion more ubiquitous in its application. On this score it is like the notions of logic: like identity, negation, alternation, and the rest. And yet, strangely, there is something logically repugnant about it. For we are baffled when we try to relate the general notion of similarity significantly to logical terms”.³²

As another sign of the importance of the conceptual level, I submit that most of scientific theorizing takes place at this level. Determining the relevant dimensions involved in the explanation of a phenomenon is a prime scientific activity. And once the conceptual space for a theory has been established, theories, in the form of *equations*, that connect the dimensions can be proposed and tested.³³

Even if the advantages of representations on the conceptual level are considerable in comparison to symbolic representations, a defender of the connectionist approach may question whether this level is really needed. Could it not be that artificial neuron networks are sufficient to solve the representational problems? After all there are several kinds of networks, e.g., Kohonen networks, where information is represented in a dimensional structure, very much like it would be represented in a conceptual space.

It is true that ANNs learn about similarities but, in general, they do so very slowly and only after excessive training. One way of making the networks more efficient is to build in structural constraints when setting up the architecture of the network. However, this often means that information about the relevant domains or other dimension generating structures are added perforce to the network. In other words, this strategy presumes the conceptual level in the very construction of the network.

A related important feature of representations in terms of conceptual spaces is that *information must be sorted into domains*. On the constructive side, it was argued above that the frame problem may be circumvented by keeping track of domain-relevant information. And on the explanatory side, one can note that when we make an observation of an object or event it is located in space and time, the object has a particular color and shape, etc. Domains are ubiquitous in descriptions of cognitive processes. Furthermore, there is ample support from neurophysiology and neuropsychology for domain-specificity in the brain.³⁴

A central feature of our cognitive mechanisms is that we assign *properties* to the objects that we observe. This functions as a way of abstracting away from redundant information about objects. Kirsch says the following about the necessity of such a mechanism:

”This capacity to predicate is absolutely central to concept-using creatures. It means that the creature is able to identify the common property which two or more objects share and to entertain the possibility that other objects also possess that property. That is, to have a concept is, among other things, to have a capacity to find an invariance across a range of contexts, and to reify that invariance so that it can be combined with other appropriate invariances.”³⁵

As has been noted above, neither symbolism nor connectionism supports domain-specificity (even though artificial neural networks in general presume that the domain of the inputs is given). It should also be noted that in many semantic theories the notion of a domain is taken for granted, but no analysis is given. The conceptual level of representation will thus provide the underpinnings for this assumption.

In a sense, an artificial neuron network can also be seen as a multi-dimensional representation, where the activity of each neuron is considered to be a dimension. This way of looking at artificial neuron networks is sometimes called the *state space* approach.^{36,37} However, when the artificial neuron network is viewed from a conceptual level, the

information is seen as represented in a small number of dimensions. Jumping between two levels of representation, one can thus say that with the aid of the processes in the artificial neuron network, a considerable *reduction of the number of dimensions* represented has taken place.³⁸ Another way of describing such a reduction of dimensions is to say that the multi-dimensional input to an artificial neuron network (or to a sensory organ) is filtered into a (small) number of domains. This results in what psychologists call a *generalisation* of the input.

Even though he is not always consistent, Smolensky writes as if the mental representation takes place only on the conceptual level:

“[C]onnectionist cognitive architecture is intrinsically two-level: semantic interpretation is carried out at the level of patterns of activity while the complete, precise, and formal account of mental processing must be carried out at the level of individual activity values and connections. Mental processes reside at a lower level of analysis than mental representations.”³⁹

Summing up the comparison between the three levels of representation, one can say that the conceptual level serves as an intermediary scale between the coarse symbolic and the fine-grained connectionist representations. Without making any ontological commitment of new entities, one can say that the conceptual dimensions “emerge” from self-organising neural systems (or in artificial neuron networks). Furthermore, in order to avoid the grounding problem, symbolic representations presume a conceptual level which provides the meanings of the symbols.

If one looks at biological cognitive systems from an evolutionary point of view, the cognition of the simplest animals can only be described, in a meaningful way, on the subconceptual level. For such systems, models based on ANNs may indeed be the most appropriate. For more advanced animals, in particular mammals and birds, it is clear that there are sophisticated mechanisms of concept formation and learning. In my opinion, these mechanisms are best modelled on the conceptual level. As regards the symbolic level, I submit that it is only in humans that we find cognition that is clearly based on symbolic representations in the sense that thinking is based on the manipulation of symbols in a rule-governed manner. It is debatable whether other primates can engage in symbolic thinking of this kind. Thus the three levels represent a rough classification of the evolution of the cognitive capacities of animals.

Connections to neuroscience

When constructing an artificial agent, the suitability of a particular form of representation cannot be decided merely on the basis of how similar problems are solved in biological agents. Nevertheless, I consider data from neuroscience and psychology to be relevant when evaluating different ways of representing information. The prime use of the theory of conceptual spaces is for representing information when constructing artificial systems, and thus the theory is not primarily empirical. Nevertheless, I believe it is possible to connect it to some theories in the neurosciences. One reason for this comparison is that the information processing involved in sensorimotor control seems to be much more fundamental for the cognitive functioning of the human brain than the processes involved in symbolic manipulations. Consequently, I see it as an advantage for the theory of conceptual spaces that it can highlight the philosophical implications of neuroscientific research in this area. The symbolic paradigm is much weaker in this respect.

A first thing to note is that the cortex abounds in topographic maps, whereby neighborhood relations at the sensory periphery are preserved in the arrangement of neurons in various “deeper” CNS regions.^{40,41} For example, one finds “retinotopic” maps in the lateral geniculate nuclei which are arranged in six layers, each layer arranged in a topographic representation of the retina; there are “somatotopic” maps representing sensory positions on the body; and there are “tonotopic” maps where the orderly mapping of neurons with sound

frequencies is preserved from the cochlea to diverse areas of the auditory cortex. Another interesting aspect of these maps is that most of them preserve the modularity of the senses, in the way that distinct types of receptor neurons are sensitive to different features of our environment and these features are kept distinct in the maps higher up in the projection system. On this point, Stein and Meredith write:

“At each successive level in the central nervous system the visual, somatosensory, and auditory representations occupy spatially distinct regions that are defined functionally and anatomically (i.e., cytoarchitecturally). At the cortical level, and in most regions of the thalamus, the domain of an individual sensory modality consists of distinct maps. The map (or maps) of a single sensory modality in, for example, primary sensory cortex is distinguished from the map in extraprimory cortex it abuts by mirror-image reversals in receptive field progressions, significant changes in receptive field properties, differences in afferent/efferent organization, and/or by specialization for different submodality characteristics. In cortex the interposition of “association” areas further segregates the representations of the different sensory modalities.”⁴²

Further support for my comparison can be gained from Gallistel, who in his excellent book on learning mechanisms in biological systems, devotes an entire chapter to “Vector spaces in the nervous system.” He writes:

“The purpose of this chapter is to review neurophysiological data supporting the hypothesis that the nervous system does in fact quite generally employ vectors to represent properties of both proximal and distal stimuli. The values of these representational vectors are physically expressed by the locations of neural activity in anatomical spaces of whose dimensions correspond to descriptive dimensions of the stimulus. The term *vector space*, which refers to the space defined by a system of coordinates, has a surprisingly literal interpretation in the nervous system. The functional architecture of many structures that process higher-level sensory inputs is such that anatomical dimensions of the structure correspond to descriptive dimensions of the stimulus. There is reason to think that this correspondence is not fortuitous; rather, it is a foundation for the nervous system’s capacity to adapt its output to the structure of the world that generates its inputs.”⁴³

COMPUTATIONAL ASPECTS

The symbolic approach to information representation is intimately connected to the classical view of computation. On this view, computations are defined by the Turing machine paradigm. According to Church’s thesis everything that can be computed with symbols can be computed on a (universal) Turing machine. However, this is not all there is to computation since the thesis is based on the assumption that all information is symbolically represented. And this is exactly what is being questioned by the connectionist and the conceptual approaches to information representation.

If we look at the methods used in the “symbol crunching” of traditional AI, a clearly dominating feature of the algorithms is that they implement some form of *rule following*. The rules can be logical axioms as in an automated theorem prover, they can be syntactic rules as in a parsing program, or they can be of a cognitively more general type as in the ACT*⁴⁴ and SOAR⁴⁵ architectures. The computational methods are very often based on some heuristics for *searching in tree-like structures*.

The computational methodology of connectionism is quite different from that of the symbolic approach. As is argued by e.g. van Gelder⁴⁶ and Smolensky,⁴⁷ connectionist systems can be seen as special cases of dynamical systems. A dynamic system can in general be described by a set of possible *states* of the system, for example all the possible combinations of activities of the neurons in an ANN, and a *dynamics* for these states, for example a set of differential equations, that describe how the system changes from one state to another. Smolensky explicitly says that connectionism

“is committed to the hypothesis that mental representations are *vectors* partially specifying the state of a dynamical system (the activities of units in a connectionist network), and that mental processes are specified by the differential equations governing the evolution of that dynamical system.

[...]

The connectionist systems I will advocate hypothesize models that are not an *implementation* but rather a *refinement* of the Classical symbolic approach; these connectionist models hypothesize a truly different cognitive architecture, to which the Classical architecture is a scientifically important approximation.”⁴⁸

The computations performed by connectionist systems are of a different nature (at least when described on the subconceptual level). It is common to point out that neural computing is *distributed* and *parallel*, instead of sequential as in a Turing machine. More important is that what goes on in a network can be described as a combination of a *fast* process of spreading of activities between the neurons and a *slow* learning process of adjusting the weights of connections between the neurons. Even though each single neuron follows a simple rule for its firing (normally a function of the sum of its inputs) and some rule for updating its weights, this is not an example of rule-following in the sense described above for the the symbolic systems.

Even though it is difficult to specify a uniform computational method for the many different kinds of artificial neuron networks, different kinds of pattern recognition or pattern transformation are important for such systems. The patterns are not determined by a set of explicit exact rules as in the symbolic systems, but rather decided by various kinds of *approximations* and *optimizations*. An important aspect is that all these decisions are made at a very *local* level - there is no central processing unit.

Turning next to computations on the conceptual level, the central mathematical notion on this level is that of a *vector*. On the representational role of vectors, Gallistel writes:

“In saying that a vector is ordered, we indicate that the position in the string matters; the string <2,7> is not equivalent to the string <7,2> Whether a set of numbers or physical quantities is a vector cannot be determined from an analysis of the set itself. This classification depends on the use made of the numbers. Loosely speaking, if the numbers enter into computational operations – for example, vector addition – that generates different outputs for different orderings of the numbers, then the input sets are vectors.”⁴⁹

Objects are represented by vectors in conceptual spaces and properties of objects are represented by regions of spaces. Consequently, the computations on the conceptual level will focus on *vector calculations*, using matrix multiplications, etc. The geometrical properties of the vectors will confer their basic representational capacities. Again, it is very unnatural to view such calculation as examples of rule following. However, in contrast to connectionist systems, the represented features are generally of a holistic rather than a distributed character.

Another methodological feature that clearly distinguishes the conceptual level from the symbolic is that relations of *similarity* will play a crucial role. Similarity between objects or properties will be represented by *distances* in spaces. Churchland and Sejnowski make a similar point, even though they write about an “activation space” in an ANN:

“An activation space will also be a similarity space, inasmuch as similar vectors will define adjacent regions in space. This means that similarity between objects represented can be reflected by similarity in their representations, that is, proximity of positions in activation space. Similarity in representations is thus not an accidental feature, but an intrinsic and systematic feature. It is, consequently, a feature exploitable in processing. Nearby vectors can cancel small differences and highlight similarities, thereby allowing the network to generalize as well as distinguish.”⁵⁰

Such a notion of a distance is difficult to model in a natural way in a symbolic system. In a connectionist system distances may appear as an emergent feature, but is hard to model on a neuronal level.

Within connectionism, the state of activities of an ANN is often represented as a vector (for example, in the quotation from Churchland and Sejnoswki above). However, this kind of vector representation is, in general, different from the one studied on the conceptual level. The activation vector of an ANN is of the same dimension as the number of neurons, while on the conceptual level, the representational space is normally of a low dimension. Furthermore, the metrics of the spaces on the conceptual level are in general simple in comparison to the very complex distance metrics that is employed by an ANN after it has been trained. On the conceptual level, the irrelevant information has been filtered out, while the activation vectors describing the state of an ANN contain a lot of noise and other redundancies from the input. This double interpretation of “vector” is a source of equivocation. However, in some of his writings, P. M. Churchland seems to be aware of the double interpretation of vectors:

”Distribution and redistribution of representations gives an informational fan-out, and this is relevant to the question of how precise representing is possible with rather coarse units. Coming at this problem from another angle, distribution means that subsets of information can be pulled out and relocated with relevant information of other representations, then to be further convolved. Consider, for example, that 3-D information is implicitly carried (one might say buried) in the output from the retinal ganglion cells. How can the information be extracted and made usable? An efficient and fast way to do this is to distribute and redistribute the information, to convolve the representations through the living matrix of synapses, until it shows up on our recording electrodes as a cell tuned to stimulus velocity in a specific direction, or to an illusory boundary, or to a human face. Topographic mapping then is a means whereby vector coding can bring to heel the problem of assembling relevant information. As Kohonen’s 1984 model showed, in a competitive learning net, the system will self organize so that nearby vectors map onto nearby points of the net, assuming that the connections are short range. That the brain avails itself of this organization is not so much a computational necessity as a wiring economy. Far from being inconsistent with topographic mapping, vector coding exploits it.”⁵¹

CONCLUSION

It is generally claimed that the symbolic and the connectionist paradigms are *incompatible*. Some of the most explicit arguments for this position have been put forward by Smolensky⁵² and Fodor and Pylyshyn.⁵³ In my opinion, the two methods are not incompatible, but they arise from modelling a cognitive system on different scales or from different perspectives. The relation between the symbolic and conceptual levels on the one hand and the connectionist level on the other hand is that connectionism deals with the “fast” behavior of a dynamic system, while the conceptual and symbolic structures may emerge as “slow” features of such a system. The upshot is that one and the same system, depending on the perspective adopted, can be seen as both an associationist mechanism and as a conceptual space which, in turn, provides a grounding for a symbolic system. Thus, by changing from one perspective to the other, conceptual representations and symbolic inferences can be seen as *emerging* from dynamic processes in a connectionist system. The pivotal point is that there is no need to distinguish between two or three kinds of systems – the different perspectives can be adopted on a single information processing system.

REFERENCES

1. A. Newell and H. Simon, Computer science as empirical inquiry: symbols and search, *CACM* 19:113-116 (1976).
2. J.A. Fodor. *Representations*, MIT Press, Cambridge, MA (1981).
3. Z. Pylyshyn, *Computation and Cognition*, Bradford Books, MIT Press, Cambridge, MA (1984).
4. J.A. Fodor and Z. Pylyshyn, Connectionism and cognitive architecture: A critical analysis, *Cognition* 28:3-71 (1988).
5. S.E. Palmer, Fundamental aspects of cognitive representation, in: *Cognition and Categorization*, E. Rosch and B. B. Lloyd, eds., Lawrence Erlbaum Associates, Hillsdale, NJ, 259-303 (1978).
6. Cf. Z. Pylyshyn (1984), p. 29: “ ... to be in a certain representational state is to have a certain symbolic expression in some part of memory.”
7. Z. Pylyshyn (1984), p. 194.
8. J. Fodor (1981), p. 230.
9. J. McCarthy and P. Hayes, Some philosophical problems from the standpoint of artificial intelligence, in: *Machine Intelligence*, Vol. 4, B. Meltzer and D. Michie, eds., Edinburgh University Press, Edinburgh (1969).
10. L.-E. Janlert, Modeling change - the frame problem, in: *The robot's dilemma: the frame problem in artificial intelligence*, Z. Pylyshyn, ed., Ablex, Norwood, NJ (1987).
11. J. Stewart, Cognition = life: implications for higher-level cognition, *Behav. Proc.* 35:311-326 (1996).
Quotation from p. 317.
12. J. Fodor (1981), p. 231.
13. J. Fodor (1981), p. 203.
14. S. Harnad, The symbol grounding problem, *Physica D* 42:335-346 (1990).
15. J. Stewart (1996), p 323.
16. C.G. Hempel. *Aspects of Scientific Explanation, and Other Essays in the Philosophy of Science*, Free Press, New York, NY (1965).
17. N. Goodman. *Fact, Fiction, and Forecast*, Harvard University Press, Cambridge, MA (1955).
18. D.E. Rumelhart and J.L. McClelland. *Parallel Distributed Processing*, Vols. 1 & 2, Bradford Books, MIT Press, Cambridge, MA (1986).
19. S.E. Palmer (1978), pp. 270-272.
20. T. Kohonen. *Self-Organization and Associative Memory*, 2nd Edition, Springer-Verlag, Berlin (1988).
21. H.R. Schiffman. *Sensation and Perception*, 2nd ed., John Wiley and Sons, New York, NY (1982).
22. P. Gärdenfors, Induction, conceptual spaces and AI, *Philosophy of Science* 57:78-95 (1990).
23. P. Gärdenfors, Frameworks for properties: Possible worlds vs. conceptual spaces, in: *Language, Knowledge, and Intentionality*, L. Haaparanta, M. Kusch, and I. Niiniluoto, eds., Helsinki (*Acta Philosophica Fennica*. 49:383-407) (1990).
24. P. Gärdenfors, Three levels of inductive inference, in: *Logic, Methodology, and Philosophy of Science IX*, D. Prawitz, B. Skyrms and D. Westerståhl, eds., Elsevier Science, Amsterdam, 427-449 (1994).
25. A. Gabriellson. Music psychology - A survey of problems and current research activities, in: *Basic Musical Functions and Musical Ability*, Publications issued by the Royal Swedish Academy of Music, 32:7-80 (1981).
26. P. Gärdenfors, Semantics, conceptual spaces and the dimensions of music, in: *Essays on the Philosophy of Music*, V. Rantala, L. Rowell, and E. Tarasti, eds., Helsinki (*Acta Philosophica Fennica*, vol. 43:9-27) (1988).
27. See H.R. Schiffman (1982), p. 138.
28. G. Lakoff. *Women, Fire, and Dangerous Things*, The University of Chicago Press, Chicago (1987).
29. R. W. Langacker. *Foundations of Cognitive Grammar*, Vol. 1, Stanford University Press, Stanford, CA (1987).
30. P. Gärdenfors (1990).
31. W.V.O. Quine, Natural kinds, in: *Ontological Relativity and Other Essays*, Columbia University Press, New York, NY, 114-138 (1969).
32. W.V.O. Quine (1969), p.117.
33. For a discussion of the role of conceptual spaces in science, see P. Gärdenfors (1990) and (1991).
34. A. Karmiloff-Smith. *Beyond Modularity: A Developmental Perspective on Cognitive Science*, Bradford Books, MIT Press, Cambridge, MA (1992). She points out that the fact that the brain organizes information in a domain-specific way does not entail that the brain is *modular* as has been argued by J. Fodor.
35. D. Kirsch, Today the earwig, tomorrow man?, *Artificial Intelligence* 47:161-184 (1991). Quotation from p. 163.
36. P. M. Churchland, Some reductive strategies in cognitive neurobiology, *Mind* 95, no. 379:279-309 (1986).
37. J. Foss, The percept and vector function theories of the brain, *Philosophy of Science* 55:511-537 (1988).

38. Neurophysiological processes are often described in this way too. For example, C.R. Gallistel. *The Organization of Learning*, Bradford Books, MIT Press, Cambridge, MA (1990) writes on p. 519: "At the retina, the infinitely dimensional spectral vector is collapsed into a three-dimensional representation by the cones. The cone representation of spectra, while it grossly reduces the dimensionality of the stimulus, at least preserves the unipolarity of light. This unipolarity is lost, however, in the coordinate transformation that occurs between the cones and V1. By the time color signals appear in V1, they are bipolar variables."
39. P. Smolensky, On the proper treatment of connectionism, *Behavioral and Brain Sciences* 11:1-23 (1988). Quotation from p. 223.
40. P.S. Churchland. *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. Bradford Books, MIT Press, Cambridge, MA (1986). Quotation from pp. 281-283.
41. In P.M. Churchland (1986), we find the following general idea: "... the brain represents various aspects of reality by a position in a suitable state space; and the brain performs computations on such representations by means of general coordinate transformations from one state space to another."
42. B.E. Stein and M.A. Meredith, *The Merging of the Senses*, MIT Press, Cambridge, MA. (1993). Quotation from p. 83.
43. C.R. Gallistel (1990), p. 477.
44. J.R. Anderson. *The Architecture of Cognition*, Harvard University Press, Cambridge, MA (1983).
45. A. Newell. *Unified Theories of Cognition*, Harvard University Press, Cambridge, MA (1990).
46. T. van Gelder, What might cognition be, if not computation?, *Journal of Philosophy* 92:345-381 (1995).
47. P. Smolensky (1988).
48. P. Smolensky, Connectionism, constituency and the language of thought, in: *Meaning in Mind: Fodor and His Critics*, B. Loewer and G. Rey, eds., Blackwell's, Oxford (1991). Quotation from p. 202-203. Also see his (1988), hypothesis 8 on p. 6.
49. C.R. Gallistel (1990), p. 476.
50. P.S. Churchland and J. Sejnowski. *The Computational Brain*, Bradford Books, MIT Press, Cambridge, MA (1992). Quotation from p. 169.
51. P. M. Churchland. *A Neurocomputational Perspective*, Bradford Books, MIT Press, Cambridge, MA (1989). Quotation from p. 171.
52. P. Smolensky (1988).
53. J. Fodor and Z. Pylyshyn (1988).

-
- ¹A. Newell and H. Simon, Computer science as empirical inquiry: symbols and search, *CACM* 19:113-116 (1976).
- ²J.A. Fodor. *Representations*, MIT Press, Cambridge, MA (1981).
- ³Z. Pylyshyn, *Computation and Cognition*, Bradford Books, MIT Press, Cambridge, MA (1984).
- ⁴J.A. Fodor and Z. Pylyshyn, Connectionism and cognitive architecture: A critical analysis, *Cognition* 28:3-71 (1988).
- ⁵S.E. Palmer, Fundamental aspects of cognitive representation, in: *Cognition and Categorization*, E. Rosch and B. B. Lloyd, eds., Lawrence Erlbaum Associates, Hillsdale, NJ, 259-303 (1978).
- ⁶Cf. Z. Pylyshyn (1984), p. 29: “ ... to be in a certain representational state is to have a certain symbolic expression in some part of memory.”
- ⁷Z. Pylyshyn (1984), p. 194.
- ⁸J. Fodor (1981), p. 230.
- ⁹J. McCarthy and P. Hayes, Some philosophical problems from the standpoint of artificial intelligence, in: *Machine Intelligence*, Vol. 4, B. Meltzer and D. Michie, eds., Edinburgh University Press, Edinburgh (1969).
- ¹⁰L.-E. Janlert, Modeling change - the frame problem, in: *The robot's dilemma: the frame problem in artificial intelligence*, Z. Pylyshyn, ed., Ablex, Norwood, NJ (1987).
- ¹¹J. Stewart, Cognition = life: implications for higher-level cognition, *Behav. Proc.* 35:311-326 (1996). Quotation from p. 317.
- ¹²J. Fodor (1981), p. 231.
- ¹³J. Fodor (1981), p. 203.
- ¹⁴S. Harnad, The symbol grounding problem, *Physica D* 42:335-346 (1990).
- ¹⁵J. Stewart (1996), p 323.
- ¹⁶C.G. Hempel. *Aspects of Scientific Explanation, and Other Essays in the Philosophy of Science*, Free Press, New York, NY (1965).
- ¹⁷N. Goodman. *Fact, Fiction, and Forecast*, Harvard University Press, Cambridge, MA (1955).
- ¹⁸D.E. Rumelhart and J.L. McClelland. *Parallel Distributed Processing*, Vols. 1 & 2, Bradford Books, MIT Press, Cambridge, MA (1986).
- ¹⁹S.E. Palmer (1978), pp. 270-272.
- ²⁰T. Kohonen. *Self-Organization and Associative Memory*, 2nd Edition, Springer-Verlag, Berlin (1988).
- ²¹H.R. Schiffman. *Sensation and Perception*, 2nd ed., John Wiley and Sons, New York, NY (1982).
- ²²P. Gärdenfors, Induction, conceptual spaces and AI, *Philosophy of Science* 57:78-95 (1990).
- ²³P. Gärdenfors, Frameworks for properties: Possible worlds vs. conceptual spaces, in: *Language, Knowledge, and Intentionality*, L. Haaparanta, M. Kusch, and I. Niiniluoto, eds., Helsinki (*Acta Philosophica Fennica*. 49:383–407) (1990).
- ²⁴P. Gärdenfors, Three levels of inductive inference, in: *Logic, Methodology, and Philosophy of Science IX*, D. Prawitz, B. Skyrms and D. Westerståhl, eds., Elsevier Science, Amsterdam, 427-449 (1994).
- ²⁵A. Gabrielsson. Music psychology - A survey of problems and current research activities, in: *Basic Musical Functions and Musical Ability*, Publications issued by the Royal Swedish Academy of Music, 32:7-80 (1981).
- ²⁶P. Gärdenfors, Semantics, conceptual spaces and the dimensions of music, in: *Essays on the Philosophy of Music*, V. Rantala, L. Rowell, and E. Tarasti, eds., Helsinki (*Acta Philosophica Fennica*, vol. 43:9-27) (1988).
- ²⁷See H.R. Schiffman (1982), p. 138.

-
- 28G. Lakoff. *Women, Fire, and Dangerous Things*, The University of Chicago Press, Chicago (1987).
- 29R. W. Langacker. *Foundations of Cognitive Grammar*, Vol. 1, Stanford University Press, Stanford, CA (1987).
- 30P. Gärdenfors (1990).
- 31W.V.O. Quine, Natural kinds, in: *Ontological Relativity and Other Essays*, Columbia University Press, New York, NY, 114–138 (1969).
- 32W.V.O. Quine (1969), p.117.
- 33For a discussion of the role of conceptual spaces in science, see P. Gärdenfors (1990) and (1991).
- 34A. Karmiloff-Smith. *Beyond Modularity: A Developmental Perspective on Cognitive Science*, Bradford Books, MIT Press, Cambridge, MA (1992). She points out that the fact that the brain organizes information in a domain-specific way does not entail that the brain is modular as has been argued by J. Fodor.
- 35D. Kirsch, Today the earwig, tomorrow man?, *Artificial Intelligence* 47:161-184 (1991). Quotation from p. 163.
- 36P. M. Churchland, Some reductive strategies in cognitive neurobiology, *Mind* 95, no. 379:279-309 (1986).
- 37J. Foss, The percept and vector function theories of the brain, *Philosophy of Science* 55:511-537 (1988).
- 38Neurophysiological processes are often described in this way too. For example, C.R. Gallistel. *The Organization of Learning*, Bradford Books, MIT Press, Cambridge, MA (1990) writes on p. 519: “At the retina, the infinitely dimensional spectral vector is collapsed into a three-dimensional representation by the cones. The cone representation of spectra, while it grossly reduces the dimensionality of the stimulus, at least preserves the unipolarity of light. This unipolarity is lost, however, in the coordinate transformation that occurs between the cones and V1. By the time color signals appear in V1, they are bipolar variables.”
- 39P. Smolensky, On the proper treatment of connectionism, *Behavioral and Brain Sciences* 11:1-23 (1988). Quotation from p. 223.
- 40P.S. Churchland. *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. Bradford Books, MIT Press, Cambridge, MA (1986). Quotation from pp. 281-283.
- 41In P.M. Churchland (1986), we find the following general idea: “... the brain represents various aspects of reality by a position in a suitable state space; and the brain performs computations on such representations by means of general coordinate transformations from one state space to another.”
- 42B.E. Stein and M.A. Meredith, *The Merging of the Senses*, MIT Press, Cambridge, MA. (1993). Quotation from p. 83.
- 43C.R. Gallistel (1990), p. 477.
- 44J.R. Anderson. *The Architecture of Cognition*, Harvard University Press, Cambridge, MA (1983).
- 45A. Newell. *Unified Theories of Cognition*, Harvard University Press, Cambridge, MA (1990).
- 46T. van Gelder, What might cognition be, if not computation?, *Journal of Philosophy* 92:345-381 (1995).
- 47P. Smolensky (1988).
- 48P. Smolensky, Connectionism, constituency and the language of thought, in: *Meaning in Mind: Fodor and His Critics*, B. Loewer and G. Rey, eds., Blackwell's, Oxford (1991). Quotation from p. 202-203. Also see his (1988), hypothesis 8 on p. 6.
- 49C.R. Gallistel (1990), p. 476.
- 50P.S. Churchland and J. Sejnowski. *The Computational Brain*, Bradford Books, MIT Press, Cambridge, MA (1992). Quotation from p. 169.

51P. M. Churchland. *A Neurocomputational Perspective*, Bradford Books, MIT Press, Cambridge, MA (1989). Quotation from p. 171.

52P. Smolensky (1988).

53J. Fodor and Z. Pylyshyn (1988).