

## INTRODUCTION. CONCEPTS: REPRESENTATIONS AND EVOLUTION

*Philip Van Loocke*

### ABSTRACT

Concepts appear in a crucial way in folk beliefs and utterances. Behind these concepts are deep-level representations. In most cases, the properties of these representations differ from the properties of folk concepts. It can be argued that the deep-level representations, and not the folk concepts, are causally active in the brain. However, this stance does not necessarily hold for all folk concepts in all circumstances, so that moderate ascriptivism seems defensible. Deep-level representations are subject to evolution. Even without knowing the very details about these representations, it is possible to formulate a theory that describes their evolution.

### 1. *Folk concepts and beliefs*

In daily life, people explain each other's behaviour in terms of concepts, opinions and attitudes. If X informs Y that she votes labour, then X presumes that Y is familiar with concepts like 'political party' and 'vote', and that Y is able to include these concepts in propositions and inferences. Now suppose that science demonstrates that, on a neural level, concepts are represented only in a very elusive and fragmentary way. Suppose that this would hold to an extent that it must be asserted that, from a neuropsychological point of view, concepts as they occur in folk psychology do not exist. Would persons like X and Y have to care about this? We can differentiate between two extreme views.

According to view A, concepts that are commonly used in folk psychology must fit scientific developments. If these developments justify the use of folk concepts, then we can continue using them. If, on the other hand, scientific insights show folk concepts to be nothing but illusions, then we should change our way of thinking, and replace folk concepts by new, scientifically justified terms. In this case, it would be

a humanistic duty to (re)educate the folk and to furnish it with scientifically justified mental tools. In special, cognitive scientists would have extend their use of scientific terms to situations they meet outside their work. Hence, according to view A, folk concepts ('mind') are a scientific kind: science must be allowed to prescribe the terms used in our daily thinking.

According to view B, it does not matter if science does or does not fit the concepts that we constantly use. Person X nor person Y has to worry about scientific progress concerning the study of concepts. They can ascribe concepts to each other even if, on a deeper level, they turn out not to use such concepts. This is the ascriptivist position (Dennett, 1987). As long as ascribing concepts to a person helps to predict the behaviour of this person, it is allowed to ascribe concepts to him. View B states that 'mind' is not a scientific kind.

Both extreme views have a number of drawbacks. If a proponent of view A makes recourse to recent science in order to evaluate folk psychology, this is usually to demonstrate that recent science is at odds with folk concepts. A. Clark (1996, this volume) points out that there is a weakness in the argument that is usually employed by defenders of this stance. This weakness is contained in the following assumption: if one tries to relate folk concepts with the representations that are proposed by science, it must be presupposed that a folk concept corresponds with a single, unequivocally definable deep structure.

However, helping a child or an adult to master a concept usually requires that different mental and motor actions are trained. For instance, the concept mastery of 'dog' includes mastery of conditions in which it is possible to approach a dog, and skills to identify conditions in which it is a good idea to run away from it. It includes knowledge on how to prepare its food, on how to educate the dog, and so on. In different situations, a subject using this concept employs different sub-skills of his mastery of the concept 'dog'. Therefore, a single folk concept like 'dog' may be linked with several deep-level representations. Hence, if it is asked if folk concepts can be mapped on scientifically justifiable representations, this question must be addressed in a more subtle way than usually is done by recent advocates of view A.

Also view B has its problems. This becomes most clearly evident in imaginistic examples. Suppose that a computer working like a giant look-up table would be able to behave in a way similar to ours. If behaviour

similar to human behaviour is a sufficient reason to ascribe concepts and beliefs, then we would have to ascribe them to the computer. But this does not fit the intended use of the terms 'concept' or 'belief'. Hence, when using such terms, a number of presuppositions about the inner workings of the believing thing are made.

Andy Clark works out an intermediate position. There is a relatively small subset of folk concepts that must be confronted with science. If it turns out that these folk concepts do not correspond with deeper level neuropsychological facts, they should be adapted. The concepts at issue relate to how cognitive systems deal with concepts. Folk psychology assumes, for instance, that people have a memory from which items can be recalled. It assumes that a cognitive system sometimes can generalize from past experience. It asserts that a belief-producer is conscious, and so on. Among such assumptions, there are assertions that are compatible with science, or assertions that can be made compatible on condition of minor modifications. Hence, even if most folk concepts would not fit deep-level representations in any perspicuous way, a small number of folk concepts about concepts remain defensible when confronted with science. Consequently, there is a core within folk psychology that should not be treated in a mere ascriptivist way. Further, given this fact, it can be defended that only systems that satisfy the properties and conditions contained in this core must be regarded as concept-manipulating or belief-holding systems. In this view, look-up tables are not to be regarded as 'true believers'.

Nevertheless, most folk-concepts may be remote from the deep level structures that have causal significance in neuro-psychological functioning. This point remains when such concepts are combined in beliefs. As Robinson (1996, this volume) points out, beliefs are properties ascribed to a total cognitive system that is more complex than this ascription suggests. It can be argued that at a problem solving brain maintains at every moment several representations active. One moment further, several of these representations may appear to be fruitless sidetracks, and may be abandoned without further elaboration. If a representation remains active longer, or if it has large causal effect, it may be called a 'dominating' representation, but it is never the only one that is active. Usually, if we ascribe a belief to a person, we make a gross simplification of the complex and composite inner dynamics of his cognitive system.

This view does not leave much room for causal influences originating

from folk concepts or beliefs. We can compare this to the noise that is produced by a car. Consider a person who does not know in any detail how a car is assembled from its parts. Suppose, however, that many years of experience with a particular car made him an expert in predicting actions to be taken when particular noises appear. Since the person does not understand the inner dynamics of the car, these actions often amount to driving the car to a garage, or to a gasoline station. Depending on the type of noise, he may try a prediction about the amount that a reparation may cost. Then, the person uses a pattern associated with a car as a whole (the noise) in order to make some plausible predictions. The noise itself, however, is not causally active. The real causal factors are hidden in inner dynamical properties. Since the person can be helped by the experience that he has acquired, he should not be discouraged to use his knowledge about predictions based on noise patterns. However, it may be useful to stimulate him to acquire some knowledge about the inner details; this would enable him to take more efficient actions. Similarly, beliefs refer to patterns that are brought forward by persons, but these patterns are not causally active. Ascribing such patterns may help to predict a person, but does not inform in any detailed way why a person is doing his deeds or thinking these thoughts.

The ascriptivist point can also be directed toward beliefs about ourselves. Hence, this argument suggests that genuine self-knowledge is hard to achieve (Nisbett and Wilson, 1977). Since self-knowledge is essential in order to allow for self-determination, insights that enhance knowledge about oneself have humanistic value. If these insights are such that they require the introduction of concepts that are not regarded as folk-concepts, this gives us an argument in favour of a complementation of folk concepts by scientific concepts. In some cases, may entail that particular folk-concepts must be abandoned.

Some of our concepts and beliefs can be characterized as 'self-evident': they are constantly used to organize our interaction with the physical and the social world, and when they are doubted, this is at most in a context of abstract philosophical exercise. For instance, we presume that other persons do have a mind. We hold the belief that, at our order of magnitude, space is Euclidean. We assume that sometimes we can rely on an elementary principle of induction, and so on. It may be uttered that, in contradistinction with other folk concepts or beliefs, such self-evident beliefs do have a causal role in our psychobiological functioning.

After all, they concern constant and very basic features behind the superficial but massive variation of stimuli in our world. Moreover, if a genetic pre-wiring of such beliefs would increase the quickness and ease in arriving at them, this might entail an increase in fitness. Therefore, Nozick (1993) suggests that such beliefs have a genetic basis. This would entail a strong argument in favour of the causal significance of a subset of folk-psychological concepts and beliefs: if particular beliefs are supported genetically, then they must have a causal role, since genetic selection only maintains features that help an organism to function in its environment.

As William Robinson (1996, this volume) points out, this argument can not save the causal significance of folk concepts. For instance, a belief in an elementary principle of induction is like an 'afterthought'. Like most other animals, humans are subject to a psychological mechanism of conditioning. This mechanism operates on a level far below the level of beliefs, and it has been selected at a stage where belief-holding animals were not yet evolved. Similarly, also animals that do not have beliefs do have some spatial insight. A human can make these insights explicit in terms of beliefs, but this does not entail that the explicitation itself has been the object of selection. It can be argued that the fitness of an organism is enhanced if some of its beliefs could be made continuous with knowledge at a pre-belief level. This means that knowledge in one system would be reencoded in another subsystem of the organisms' brain. In present literature, this theme of representational redescription is encountered frequently in order to formulate some of the core ideas of ontogenetic psychology (Karmiloff-Smith, 1992).

To the extent that a such a representational redescription sometimes accurately leads to beliefs, the latter may be 'reliable' reflections or abstractions of principles operating at a lower level; but the processes redescribed remain the causally effective ones. However, stating that a limited set of beliefs may be reliable in this sense is a point against an eliminativism that would be too strong. Further, also these self-evident concepts and beliefs can be confronted with scientific insights. Sometimes, the latter are at odds with the beliefs, like in case of relativity theory and quantum theory. Since such scientific insights typically have large bearing on the worldview of people, and since allowing people to construct worldviews that are tenable in the light of current science is a respectable aim, it can be defended that also such folk concepts and

beliefs should be adapted. We notice that this argument, along with the argument of self-cognition, is basically of an ethical nature. Hence, the epistemological argument in favour of the stance that some folk concepts must be adapted to science can be complemented with strong ethical arguments.

## *2. The evolution of concepts.*

If, nevertheless, most of our concepts and beliefs are to be interpreted in an ascriptivist sense, then models about psychological functioning should not attribute causal roles to folk concepts. Hence, terms referring to concepts should not be the central in these theories. However, for a given concept or set of concepts, one can often plausibly argue that particular relational properties or higher-order characteristics must be invariant over different subjects, and that they must play a causal role. For instance, someone's knowledge about a restaurant includes information about sequences of behaviour typically shown in a restaurant (Shank and Abelson, 1977). In a connectionist context, it has been argued that a person disposes of a 'restaurant-schema' that allows him to interpret a relatively wide range of typical situations he may encounter at such a place (Rumelhart, Smolensky et al., 1986). Such a schema may be represented in a fairly distributed way, and the set of units involved in its representation may be subject to fluctuations. If subsets of such units can be attributed meaning, then it is often at a level that is 'sub-semantic' relatively to the level of folk concepts (Smolensky, 1988). Nor schema's, nor sub-semantic features are folk-concepts. From the perspective of a schema-model, a concept can usually be associated with a collection of units that is a relatively variable subset of a schema, or with variable sets of units belonging to different schema's.

For the present discussion, however, it matters that it can be scientifically conjectured that there are structures that represent information, such a schema's. In some cases, it can be plausibly argued with help of experimental psychology what such representations must be able to accomplish. Therefore, it remains meaningful to develop causal theories about the higher-order structures behind folk-concepts (Kitcher, 1987). In special, one may be interested in the evolution of these structures. Biological evolution theory offers a remarkable precedent. Darwin did not

know about DNA's, or chromosomes. Nevertheless, his theory has had tremendous importance for science during the last two centuries. With this precedent in mind, a growing number of philosophers and cognitive scientists try to found a science of conceptual evolution. The fact that the nature of representations is not yet understood in a detailed way does not prohibit the formulation of a theory that explains some properties of the phenotypes of these representations.

One place to look for these phenotypes is in the domain of beliefs that are communicated. As we have seen, beliefs can be compared to patterns. If a subject that is addressed in communication consciously or subconsciously recognizes a belief-pattern, this recognition may be reflected in his long term memory and affect representations that are causally active. It can be noticed that an evolution theory for concepts draws to some extent beliefs in terms of folk concepts into the scientific theory, namely by giving them the role of phenotypes.

Liane Gabora (1996, this volume) examines the conditions that have to be present in order to be able to speak of evolution on the level of representations. First, it is necessary that actual representations belong to a large space of possible representations. Second, there must be ways to generate variations of representations. Third, there must be a procedure to select some representations over others. Such a procedure may be based on a fitness landscape that is defined over the space of possible representations. Finally, there must be a way of replicating and transmitting representations in which variations have been included.

Since it can be argued that these conditions are fulfilled for representations, it is meaningful to develop an evolutionary theory for concepts and beliefs. It is to be expected that conceptual evolution parallels genetic evolution in some respects, but also that it differs from it in a number of other ways. One of them is that variations are less random than in case of biology. Representations that have been subject to variation have a greater probability than chance of being more fit than their predecessor.

The main reason for this is that the vehicle that varies and reproduces representations contains information about the fitness landscape. The question if a representation is fit or not depends on the physical and social environment in which a person has to function. His brain, however, makes maps and mental models of these environments. At this instance, it should be noticed that, unlike genes, representations do not include instructions for their replication; they must rely on brains to do

it for them. Hence, the vehicle for replication (the brain) contains an interior model of the external fitness landscape. This interiorized landscape is used when representations are made subject to variation. Therefore, variations that lead to fitter representations have higher probability to occur.

Instead of 'representation', one frequently uses the term 'meme' to refer to the entity that is evolved in the conceptual sphere. Also Plotkin's contribution (1996, this volume) aims at clarifying the term 'meme', and it introduces a hierarchy of more and less fundamental memes. Historically, the terms 'meme' and 'meme-pools' (Dawkins, 1976) or culture-gens (Lumsden and Wilson, 1981) have been proposed as cultural analogues of genes and gene pools. Dawkins (1976) defined memes as 'units of imitation', and proposed that they are exemplified by 'tunes, ideas, catch phrases, ways of making pots or of building arches'. In agreement with the point that is made in this chapter, attempts to make this notion more specific suggest that actions and artefacts may not be regarded as real instances of memes. Even verbal acts that express beliefs are patterns that correspond to vehicles of selection (or phenotypes) rather than genotypes, which are representations at a deeper level (Heyes and Plotkin, 1989).

Memes must be searched in structures that reside in memories of individuals, and that use actions and communication for their dissemination. This suggestion, however, can not be interpreted at the deepest neurological level, since at this level, different persons represent the same schema's or knowledge structures in profoundly different ways. Then, the representations to look at must be definable at a more macroscopic, or at a more intermediate level. This agrees with a connectionist proposal according to which representations like schema's have to be situated at such an intermediate level (Smolensky, 1988).

In continuity with the present discussion, Henry Plotkin suggests that not individual concepts, but more abstract schema's or higher order structures figure as important types of memes. For instance, a concept that relates to a concrete school is not meme; a better meme-candidate is the higher knowledge structure that specifies that schools are places where children go, where they learn to read and write, where they are taught by teachers, and so on. Around such stable memes, clouds of large numbers of less stable and more superficial memes may cluster. Below the relatively stable cultural memes are a number of culturally-universal meme structures. These are of two kinds. First, there are meme-struc-



tures that are the causally effective representations associated with 'self-evident beliefs' of the type discussed by William Robinson (this volume), such as the belief that other persons have a mind. Second, some deep meme-structures work largely outside the focus of consciousness. An instance of such a meme is the 'social force' that stimulates the participation of humans in groups and that supports consensus formation (Sherif, 1936; Jacobs and Campbell, 1961).

If representations are made subject to processes that have an analogy in biological evolution, then current connectionist models (such as the schema model) are facing a problem. Among the methods of artificial intelligence, genetic algorithms have more affinity with genetic evolution than have connectionist models. One remarkable difference between both methods is that a genetic algorithm solves a problem by considering a large population of representations at once, and by letting evolve these representations until one of them gives a solution for the problem at hand. A neural network, on the other hand, considers a single distributed representation, and tries to transform this single representation gradually into a solution by a successive series of neural updates. In my contribution to this volume, I show how this limitation can be overcome. I explain the connectionist model QNET that has units that are active in different frequencies, and briefly recapitulate its properties. Then, I show how it can be used in conjunction with genetic algorithms so that both the genetic as well as the connectionist procedure can become more powerful. Hence, I argue that evolutionary explanations are compatible with connectionist ones.

In the final chapter of this issue, Markus Peschl considers scientific concepts. Studying the evolution of scientific concepts has the advantage that the relations between different concepts have been fleshed out and made as explicit as possible in the process of scientific investigation. Peschl argues that scientific concepts can nevertheless be considered from the same evolutionary perspective as pre-scientific ones. The general epistemological view that is suggested by such an approach is a constructivist one. An organism develops a representation of its environment in order to increase its fitness, and not in order to create a mirror for the environment. There are significant respects in which this property remains when highly abstract scientific concepts are developed. It is not to be expected from this perspective that biological systems develop a propositional code with a nicely organized and perspicuous semantics.

A theory, as a construct of a nervous system, can be considered as a point in a special instance of a weight space, called a theory space. Scientific activities during episodes of 'normal' science can be described as searching the weight space for points that correspond to better fitting theories. In episodes of new paradigm construction, the weight space itself is varied according to a process that can be compared with variations in the genetically defined characteristics of the nervous system itself.

### 3. *Concepts and philosophical conceptualism*

Genetic algorithms always refer to a fitness landscape. The peaks of such a landscape correspond to the attractors of the evolving entities, whether these are genes or memes. Here 'attractor' means that an entity that is relatively close to it will evolve in such a way as to coincide with the attracting state. The term 'attractor' is nowadays frequently encountered in several branches of what may be called the sciences of complexity, such as chaos theory, fractal theory, algorithmic complexity theory, neural network theory, and so on. The extent to which descriptions in such sciences are replete with attractors has some relevance for the issue of philosophical conceptualism.

Let us remind of what philosophical conceptualism is about. Basically, there are two conceptualist stances that must be differentiated. The first one states that there is more to a concept than a symbol that appears in a syntactically ordered network of symbols. In addition to representations that are syntactically organized, concepts are represented also in ways that are continuous with mental imagery, in representations with a topological organization, in formats that are used in mental models, and so on. In other terms, there is more to the representation of a concept than some deep-structure linguistic code, like some nominalistic approaches would want. This version of conceptualism receives much support these days (Donald, 1991).

The fact that memes evolve on an attractor landscape is relevant for a stronger version of conceptualism. According to this version, there is a respect in which the existence of memes surpasses the individual brains in which they are physically realized. An extreme version of this stance

is platonism, that asserts that (particular) concepts exist in a world that is at most equally real as our experiential world. Such a stance does not receive support from the present considerations. In order to see how a more moderate stance can get some support, suppose that, given an environment that satisfies certain conditions, the following condition C holds:

C: Fitness landscapes associated with different brains have in some regions peeks at comparable places

In that case, these peeks will be inter-subjective attractors for memes. Some of these attractors may not yet have attracted any meme, for instance because memes generated thus far were too far removed from their neighbourhood. Strong attractors, however, are to be regarded as potential memes. In this sense, a potential meme may be said to exist before a single human ever realized a thought in terms of this meme. This existence, however, is only an existence as an attractor state. Most memes are not attractors in this sense, since they are strongly biased by the experiential world of the particular person that holds them. Furthermore, some parts of the fitness landscape may be too dependent on a concrete feature of the (varying) environment in order to contain attractors that last over an extended region of time. But maybe deep memes (Plotkin, this volume), or concepts corresponding to natural numbers may be regarded as such attractors.

Condition C refers to 'some regions' in the fitness landscape, and it is formulated on the presumption that the environment satisfies 'certain' conditions. Surely this type of discussion will have to be cast in more precise terms in order to be of use for more deep philosophical issues. The contributions to the present volume are just one step in the conceptual adventure that consists of the clarification of the evolution of concepts.

Universiteit Gent

*Acknowledgement:* This issue is a result of the project 'Concepts: representation and evolution' of the lab for Applied Epistemology at the University of Ghent. The editor was supported by a grant of the Flemish Fund for Scientific Research. Participants in this project, in alphabetical order, were: Michael Arbib, Andy Clark, Terry Darnall, Liane Gabora, Ben Goertzel, Yozef Grodzinsky, Mark Hubey, Markus Peschl, Steven Pinker, Henry Plotkin, William Robinson, Lee Spector, Paul Thagard, Wolfgang Tschacher, Philip Van Loocke and Lucia Vaina. The editor wishes to thank all participants for their effort and

for many fruitful discussions. Contributions not included in this issue appear in *Communication and Cognition* (1996, 29(3)) and in *C&C-ai* (1996, 13(4)).

## REFERENCES

- Clark A. (1996), Is mind a scientific kind?, *this volume*.
- Dawkins R. (1976), *The selfish gene*, Oxford: Oxford University Press.
- Donald M. (1991), *Origins of the modern mind*, Harvard University Press.
- Gabora L. (1996), A day in the life of a meme, *this volume*.
- Heyes C., Plotkin, H. (1989), Replicators and interactors in cultural evolution, in M. Ruse (ed.), *What philosophy of biology is*, Dordrecht: Kluwer, pp. 139-162.
- Jacobs R., Campbell D. (1961), The perpetuation of an arbitrary tradition through several generations of a laboratory microculture, *Journal of Abnormal and Social Psychology* **62**, 649-658.
- Karmiloff-Smith A. (1992), *Beyond modularity: a developmental perspective on cognitive science*: MIT Press.
- Kitcher P. (1987), Confessions of a curmudgeon, *Behavioral and Brain Sciences* **10**, 89-97.
- Lumsden C., Wilson, E. (1981), *Genes, mind and culture*, Cambridge: Harvard University Press.
- Nisbett R., Wilson T., Telling more than we can know: verbal reports on mental processes, *Psychological Review* **84**, 231-259.
- Peschl M. (1996), The development of scientific concepts and their embodiment in the representational activities of cognitive systems, *this volume*.
- Plotkin H. (1996), Some psychological mechanisms of culture, *this volume*.
- Robinson W. (1996), Evolution and self-evidence, *this volume*.
- Rumelhart D., Smolensky P., Hinton G. (1986), Schemata and sequential thought, in: D. Rumelhart, J. Mc Clelland (eds.), *Parallel distributed processing*, vol. 2, Cambridge: Bradford Books/MIT Press.
- Shank R., Abelson R. (1977), *Scripts, plans, goals and understanding*, Hillsdale NJ: Erlbaum.
- Sherif M. (1936), *The psychology of social norms*, New York: Harper and Row.
- Smolensky P. (1988), On the proper treatment of connectionism, *Behavioral and Brain Sciences* **11**, 1-74.
- Van Loocke Ph. (1996), The connectionist model QNET and its combination with genetic algorithms, *this volume*.