published as:
Heylighen F. (1993): "Selection Criteria for the Evolution of Knowledge", in: Proc. 13th Int. Congress on Cybernetics (Association Internat. de Cybernétique, Namur), p. 524-528.

Selection Criteria for the Evolution of Knowledge

Francis HEYLIGHEN*

PO, Free University of Brussels, Pleinlaan 2, B-1050 Brussels, Belgium

E-mail: fheyligh@vnet3.vub.ac.be

Introduction

One of the essential issues of the Principia Cybernetica Project, which aims at the development of an evolutionary-cybernetic philosophy (Turchin, 1991; Heylighen, 1991ab), is *epistemology*, or the theory of knowledge. When we look at the history of epistemology, we can discern a clear trend, in spite of the confusion of many seemingly contradictory positions. The first theories of knowledge, such as Platonic idealism, empiricism and the reflection-correspondence theory stressed its absolute, passive and permanent character. These approaches try to formulate unambiguous, fixed criteria for distinguishing "true" or "real" knowledge from "false" one. Later theories, starting with conventionalism, pragmatism and up to constructivism, put the emphasis on the relativity or situation-dependence of knowledge, its continuous development or evolution, and its active interference with the world and its subjects and objects. Here the criteria, such as problem-solving competence, coherence and consensus are more context-dependent and variable.

A more synthetic outlook is offered by *evolutionary epistemology* (Campbell, 1974). Here it is assumed that knowledge is constructed by the subject or group of subjects in order to adapt to their environment in the broad sense. Construction happens through blind variation of existing pieces of knowledge, and the selective retention of those new combinations that somehow contribute most to the survival and reproduction of the subject(s) within their given environment. Multiple criteria, biological, cognitive as well as social, determine which knowledge survives that ongoing process of natural selection.

A most recent, and perhaps most radical approach, extends this evolutionary view in order to make knowledge actively pursue goals of its own. This approach, which as yet has not had the time to develop a proper epistemology, may be called *memetics* (Dawkins, 1976; Moritz, 1991; Heylighen, 1992a). It notes that knowledge can be transmitted from one subject to another, and thereby loses its dependence on any single individual. A piece of knowledge that can be transmitted or replicated in such a way is called a "meme". The death of an individual carrying a certain meme now no longer implies the elimination of that piece of knowledge, as evolutionary epistemology would assume. As long as a meme spreads more quickly to new carriers, than that its carriers die, the meme will proliferate, even though the knowledge it induces in any individual carrier may be wholly inadequate and even dangerous to survival. In this view a piece of knowledge may be successful (in the sense that it is common or has many carriers)

* Senior Research Assistant NFWO (Belgian National Fund for Scientific Research)

even though its predictions may be totally wrong, as long as it is sufficiently "convincing" to new carriers.

Though such a view may appear to lead to pure relativism, the present paper will argue that, in spite of the variability and subjectivity of knowledge, a number of clear, though sometimes inconsistent, criteria can be formulated, that determine the selection of adequate knowledge and the elimination of inadequate knowledge. The more criteria a piece of knowledge fulfils, and the stronger the fulfilment, the more likely that the piece of knowledge will maintain and proliferate.

The Function of Knowledge

In order to derive criteria for the selection of knowledge we must first make clear what knowledge is, and what it is used for. In an evolutionary-cybernetic philosophy (Heylighen, 1991a,b), knowledge is seen as *that which allows a control system to select the actions that will make its survival and reproduction more likely*, in a given environment, offering different resources, and exerting different types of perturbations on the system. Adequate selection of actions requires an (explicit or implicit) anticipation of the likely effects of perturbations and actions, so that actions may be chosen that will counteract the negative effects of perturbations *before* those have become so strong as to endanger the survival of the system. Hence, the essential function of knowledge is *prediction* (Turchin, 1991).

The structure of an elementary prediction or piece of knowledge can be represented as a *production rule* "if a certain phenomenon *A* is distinguished, then, either perform a certain action *B*, or expect a certain other phenomenon *B*" (Holland et al., 1986). With symbols: $A \rightarrow B$, where *A* and *B* are *distinctions* (Heylighen, 1989; 1992c), discriminating between the presence and the absence of certain types of phenomena or actions. Cognitive systems are built up as sets of interconnected production rules. Two rules are connected if the output condition of one rule matches the input condition of another rule.

Perhaps the most fundamental question that any theory of knowledge or cognition must answer is how an infinitely complex environment can be modelled, allowing predictions, by a system that is necessarily much simpler (contains less distinctions) than this environment. One part of the answer comes from the above cybernetic view of knowledge: phenomena in the environment only need to be modelled insofar that they represent a potential threat to the survival of the system (that is, a perturbation from its homeostatic functioning), or an opportunity for growth and reproduction. All knowledge is in this sense fundamentally subjective, focused (directly or indirectly) on the 'selfish' (Heylighen, 1992a) purposes of the knowing system. All phenomena irrelevant to those purposes will in general be ignored, except insofar that there is a certain probability that they may become relevant in a later stage. This allows a first, enormous reduction of complexity of the model of the environment the system will have.

A second part of the answer comes from an evolutionary interpretation of the environment itself. Natural selection privileges systems with an invariant or stable organization (Heylighen, 1992b). Most systems will, hence, be relatively constant, and the number of their features that undergo changes at any given time will be much smaller than the number of features that remain the same. For example, when leaning against a tree or a wall, we normally do not expect the wall to move, crumble or change shape. Since invariant phenomena normally cannot be controlled, this means that the control system can again ignore a very large number of features of its environment.

The conclusion is that the control system will normally only pay attention to phenomena that: 1) constitute some variation, transition or deviation from the normal, stable configuration; 2) may influence the system's maintenance or growth. By *default* all features that are not perceived as varying, will be assumed to remain stable. When a phenomenon fulfils the two conditions above, but there is not any rule available to make relevant predictions, a blind-variation-and-selective-retention of new rules will be triggered (Campbell, 1974). A basic triggering condition for rule generation is called the "unusualness heuristic" by Holland et al. (1986): unexpected or unusual (i.e. deviating from the normal configuration) events signal other unexpected events.

Individual Fitness Criteria

The first criteria that determine whether a newly generated production rule, or potential piece of knowledge, will be retained depend on whether the rule will increase the fitness of the individual control system or organism. As noted above, a distinction is useful only insofar that it can be used to predict some deviation or variation from equilibrium. This entails two separate characteristics:

1) some change must follow or precede the appearance of the distinguished phenomenon; phenomena that do not make any difference, are not informative, and, hence, are not considered to be real in any practical sense. This corresponds to Kelley's (1967) *distinctiveness* criterion for distinguishing "real" from "illusory" perceptions.

2) The change following or preceding the phenomenon must not be unique, but share some properties with changes associated with similar phenomena. A minimal regularity or *invariance* of effect is needed in order to make predictions (Heylighen, 1989). The invariance criterion can be subdivided in a number of more specific criteria, specifying under which transformations of initial phenomena (causes) the characteristics of the effect will be invariant. The simplest type of invariance is probably that over *time* ("consistency" according to Kelley, 1967): subsequent appearances of the distinguished phenomenon should lead to similar effects. An effect can also be invariant over *settings* or circumstances, or over different *points of view* or *modalities of perception*. The more invariant the causal relationship, the more generally reliable and applicable it is, the more predictions can be made with it, and the more useful the corresponding piece of knowledge is.

The two criteria together can be summarized by the concept of *distinction invariance* or *distinction conservation* (Heylighen, 1989), characterizing causality of processes. Such causal criteria are universal or 'objective', checking the covariation between causes and effects without distinguishing between different types of effects.

The third type of criteria entailed by the above view of knowledge will be subjective, paying special attention to the effects on the organism. Here we can distinguish between *survival criteria*, selecting rules useful for monitoring and controlling the basic variables determining whether or not the organism can survive (e.g. level of sugar in the blood, temperature, water, absence of predators...), and *reproductive criteria*, selecting rules that tell the organism how to find a mate and successfully produce offspring. The latter rules are relevant to the selection of knowledge only insofar that knowledge is passed on to the offspring (by inheritance or education).

The success of a rule will not only depend on its intrinsic adequacy for furthering the organism's fitness, though: the rule should first get the chance to get sufficiently well-established to prove its adequacy. This is the criterion of *learnability*: rules that are difficult to assimilate may never even get to the stage where they are subjected to selection governed by the other criteria. Learnability implies among other things that the rules should not be too complex, and should not too directly contradict already established rules, because of the tendency to avoid cognitive dissonance. The latter demand for a minimal consistency between old and new rules corresponds to the criterion of *coherence*, used in logical and constructivist epistemologies.

Social Fitness Criteria

Once we reach the stage where the isolated individual is replaced by a social system with communication of knowledge between individuals, knowledge will not only be selected on the basis of its capacity to maintain (within the memory of an individual), but also on the basis of its capacity to reproduce (being passed from one individual to another one). Knowledge fragments or systems of production rules that have this capacity for replication are called *memes*. Meme spreading, to be effective, must fulfil a number of 'social' criteria, that are in a number of ways different from, and sometimes even inconsistent with, the 'individual' criteria above (Heylighen, 1992a).

The first condition for a meme to spread is that the knowledge possessed by one individual must be exteriorized or *expressed*, so that another individual can again interiorize or assimilate (learn, understand, interpret) that expression. The simplest way to do that is by *imitation*: the meme (set of rules) expresses itself through the behavior of the individual applying the rules. Another individual may then watch that behavior, and try to act in a similar way. If he succeeds to mimic the behavior, we may conclude that he has interiorized the rule(s). In the case of imitation, the primary selection criterion is that a rule should lead to a *salient*, i.e. an easily perceived and remembered, behavior. Bird songs form a good example of such behavioral patterns, that are transferred between individuals by imitation. A second criterion is that the individual should be motivated to imitate that particular behavior, e.g. because it leads to clear advantages for the one carrying out the behavior.

A more effective way of communicating knowledge is by means of an intersubjective code or language, that expresses distinctions and their connections through symbols. The primary criterion here is that the expression of a production rule would faithfully capture the underlying distinctions. This requires that different individuals in different contexts should interpret the expression in the same way. This may be called the criterion of *formality* (Heylighen, 1993). Formality is less important when the contexts and individuals are relatively similar, because then the context and the already existing knowledge aid in the interpretation of the expression. Imitation could be seen as a maximally informal way of communicating knowledge, since the meaning of behavior can only be found by carrying it out in the appropriate context.

The following criteria are related to the criterion stating that a rule should lead to salient behavior: 1) the rule should be *easy to express*, and 2) the individual should be *motivated to express* it. Secret knowledge, that cannot be expressed, does not qualify for memetic replication. The first condition depends on the language, making distinctions easy to express or not, and on the intrinsic variability of the distinction, but that is already encompassed in our criterion of invariance. The second condition represents a factor of "*contagiousness*" of the meme: some memes provoke behavior in their carriers that tends to persuade or convert others to adopt the meme. For example, most religious memes include such proselytizing or propagation.

Two closely connected selection criteria for the spreading of ideas are *conformity* and *consensus*. The more people already agree upon or share a particular idea, the more easily a newcomer will in turn be infected by the meme. This can be explained by the fact that the newcomer will be subjected to expressions of the meme more often, and will more likely get in trouble if he expresses dissonant memes (Heylighen, 1992a). Though conformity pressure is mostly irrational, often rejecting knowledge that is adequate because it contradicts already established beliefs, consensus is a criterion of the invariance type, since it implies that a belief does not vary over individuals. In the section on individual criteria, it was stated that more invariant rules are more reliable predictors, and consensus, in the sense where people agree about an idea because they independently came to the same conclusion, can be viewed as a quite rational criterion for selecting knowledge (cf. Kelley, 1967).

Conclusion

The complete set of selection criteria that we have enumerated should provide a good checklist for determining the chances that a particular production rule or knowledge fragment would maintain and proliferate. By way of illustration, we may apply them to a typical “good idea”, the obviously successful discovery of the wheel. Using or not using wheels leads to clearly *distinct* results. Those results are quite stable, *invariant* over time, ways of observing, or circumstances. Making and using wheels is *easy to learn*, and is not *incoherent* with previously established knowledge. It has many advantages that may contribute to increased chances of *survival*, e.g. the capacity to transport food and sheltering materials over large distances. Indirectly, that also makes it easier for groups using wheels to *reproduce* and spread. The use of wheels is quite noticeable or *salient*, and hence easy to imitate. The structure and functioning of a wheel can also be expressed *formally*, e.g. by means of geometry and mechanics, so that the idea can be unambiguously transmitted even to people who have never actually seen a wheel. Even without formal representation, the concept of a wheel is relatively easy *to express* in natural language or by demonstration. People who are enthusiastic about the advantages of using wheels will in general also be *motivated* to convince others about the usefulness of the idea. And once the idea has spread, most people will rely on wheel-driven vehicles, so that newcomers will tend to quickly *conform* to that consensual pattern of behavior.

References

Campbell D.T. (1974): "Evolutionary Epistemology", in: *The Philosophy of Karl Popper,* Schilpp P.A. (ed.), (Open Court Publish., La Salle, Ill.), p. 413-463.

Dawkins R. (1976): *The Selfish Gene*, (Oxford University Press, New York).

Heylighen F. (1989): "Causality as Distinction Conservation: a theory of predictability, reversibility and time order", *Cybernetics and Systems* 20, p. 361-384.

____ (1991a): "Cognitive Levels of Evolution: pre-rational to meta-rational", in: *The Cybernetics of Complex Systems - Self-organization, Evolution and Social Change*, F. Geyer (ed.), (Intersystems, Salinas, California), p. 75-91.

____ (1991b): "Evolutionary Foundations for Metaphysics, Epistemology and Ethics", in : *Workbook of the 1st Principia Cybernetica Workshop*, Heylighen F. (ed.) (Principia Cybernetica, Brussels-New York), p. 33-39.

____ (1992a) : "‘Selfish’ Memes and the Evolution of Cooperation", *Journal of Ideas*, V. 2 #4, p. 77-84.

____ (1992b): "Principles of Systems and Cybernetics: an evolutionary perspective", in: *Cybernetics and Systems '92*, R. Trappl (ed.), (World Science, Singapore), p. 3-10.

____ (1992c): "Non-Rational Cognitive Processes as Changes of Distinctions", in: *New Perspectives on Cybernetics. Self-Organization, Autonomy and Connectionism*, G. Van de Vijver (ed.), (Synthese Library v. 220, Kluwer Academic, Dordrecht), p. 77-94.

____ (1993): "Making Thoughts Explicit: advantages and drawbacks of formal expression", submitted to *Journal of Applied Philosophy*

Holland J.H., Holyoak K.J., Nisbett R.E. & Thagard P.R. (1986): *Induction : processes of inference, learning and discovery*, (MIT Press, Massachusetts).

Kelley H.H. (1967): "Attribution Theory in Social Psychology", *Nebraska Symposium on Motivation* 15, p. 192-238.

Moritz E. (1990): "Memetic Science: I - General Introduction", *Journal of Ideas* 1, p. 1-23.

Turchin V. (1991): "Cybernetics and Philosophy", in: *The Cybernetics of Complex Systems - Self-organization, Evolution and Social Change*, F. Geyer (ed.), (Intersystems, Salinas, CA), p. 61-74.

	masonry	washing	counting	Quant.Mech.	miniskirt	God
distinctive-ness		*	*	*	*	-
invariance		*	*	*		
learnability	*	*	*		*	*
survival	*	*	*	*		
reproduction					*	
salience		*		-	*	
formality			*	*		*
ease of ex-pression		*	*	-	*	*
conta-giousness					*	*
consensus		*	*	*		*
	2	7	7	5	6	5