CONCEPTS: *the elements from which propositional thought is constructed, thus providing a means of understanding the world.*

Concepts are used to interpret our current experience by classifying it as being of a particular kind, and hence relating it to prior knowledge. The concept of "concept" is central to many of the cognitive sciences. In cognitive psychology, conceptual or semantic encoding effects occur in a wide range of phenomena in PERCEPTION, ATTENTION, LANGUAGE COMPREHENSION and MEMORY. Concepts are also fundamental to REASONING, in both machine systems and people. In AI, concepts are the symbolic elements from which KNOWLEDGE REPRESENTATION systems are built in order to provide machine-based expertise. Concepts are also often assumed to form the basis for the MEANING of nouns, verbs and adjectives, (see COGNITIVE LINGUISTICS, SEMANTICS) . In behaviorist psychology, a concept is the propensity of an organism to respond differentially to a class of stimuli (for example a pigeon may peck a red key for food, ignoring other colors.) In cultural anthropology, concepts play a central role in constituting the individuality of each social group. In comparing philosophy and psychology, it is necessary to distinguish philosophical concepts understood as abstractions, independent of individual minds, and psychological concepts understood as component parts of MENTAL REPRESENTATIONS of the world (see INDIVIDUALISM).

Philosophical foundations

Philosophy distinguishes NARROW conceptual CONTENT, which is the meaning of a concept in an individual's mental representation of the world, from BROAD CONTENT in which the meaning of a concept is also partly determined by factors in the external world. There has been much debate on the question of how to individuate the contents of different concepts, and whether this is possible purely in terms of narrow content (Fodor, 1983; Kripke, 1972; Putnam, 1975). A related problem is how concepts as purely internal symbols in the mind come to stand in a symbolic relation to classes of entities in the external world.

Concepts are considered to serve two functions, an intensional and an extensional role (Frege, 1952). There are different technical ways to approach this distinction. One philosophical definition is that the extension is the set of all objects in the actual world which fall under the concept, whereas the intension is the set of objects that fall under the concept in all possible worlds. In cognitive science a less strict notion of intension has been operationalised as the set of propositional truths associated with a proper understanding of the concept -- for example that chairs are for sitting on. It resembles a dictionary definition, in that each concept is defined by its relation to others. Intensions permit inferences to be drawn, as in "This is a chair, therefore it can be sat upon", although, as the example illustrates, these inferences may be fallible. The extension of a concept is the class of objects, actions or situations in the actual external world which the concept represents and to which the concept term therefore refers (Frege's "reference"). Frege argued that intension determines extension; thus the extension is the class of things in the world for which the intension is a true description. This notion of concepts leads to a research program for the

analysis of relevant concepts, (such as "moral" or "lie") in which proposed intensional analyses of concepts are tested against intuitions of the extension of the concept, either real or hypothetical. Fodor (1994) has advanced arguments against this program. To avoid the circularity found in dictionaries, the intension of a concept must be expressed in terms of more basic concepts (the SYMBOL GROUNDING PROBLEM in cognitive science). The problems involved in grounding concepts have led Fodor to propose a strongly INNATIST account of concept acquisition, according to which all simple concepts form un-analyzable units, inherited as part of the structure of the brain. Others have explored ways to ground concepts in more basic perceptual symbolic elements (Barsalou, 1993).

Psychology and concepts

There are three main research traditions in the psychology of concepts. First, the cognitive developmental tradition, pioneered by Piaget (1967), seeks to describe the ages and stages in the growing conceptual understanding of children. Concepts are SCHEMAS. Through self- directed action and experience the assimilation of novel experiences or situations to a schema leads to corresponding accommodation of the schema to the experience, and hence to conceptual development. Piaget's theory of adult intelligence has been widely criticized for over-estimating the cognitive capacities of most adults. His claims about the lack of conceptual understanding in young children have also been challenged in the literature on conceptual development (Carey, 1985; Keil, 1989). Research in this tradition has also had a major influence on theories of adult concepts developed within the lexical semantics tradition.

The second research tradition derives from behaviorist psychology. For this tradition, concepts involve the ability to classify the world into categories (see also MACHINE LEARNING). Animal discrimination learning paradigms have been used to explore how people learn and represent new concepts. A typical experiment involves a controlled stimulus set, usually composed of arbitrary and meaningless elements, such as line segments, geometric symbols, or letters, which has to be classified into two or more classes. The stimuli in the set are created by manipulating values on a number of stimulus dimensions (for example shape or color). A particular value on a particular dimension constitutes a stimulus feature. The distribution of stimuli across the classes to be learned constitutes the structure of the concept. Experiments typically involve training involving trial and error learning with feedback. In a subsequent transfer or generalization phase, novel stimuli are presented for classification without feedback, to test what has been learned. Three types of model have been explored in this paradigm. Rule-based learning models propose that participants try to form hypotheses which are consistent with the feedback in the learning trials (see for example Bruner Goodnow and Austin, 1956). Prototype learning models propose that participants form representations of the average or prototypical stimulus for each class, and perform the classification by judging how similar the new stimulus is to each prototype. Exemplar models propose that individual exemplars and their classification are stored in memory, and that classification is based on the relative average similarity of a stimulus to the stored exemplars in each class,

usually assuming an exponential decay of similarity as distance along stimulus dimensions increases (Nosofsky, 1988). Exemplar models typically provide the best fits to experimental data, although rules and prototypes may also be used when the experimental conditions are favorable to their formation. NEURAL NETWORK models of category learning capture the properties of both prototype and exemplar models, since they abstract away from individual exemplar representations, but at the same time are sensitive to patterns of co-occurrence of particular stimulus features.

The study of categorization learning in the behaviorist tradition has generated powerful models of fundamental learning processes with an increasing range of application. As yet however the connection to other traditions in the psychology of concepts (for example cognitive development or lexical semantics) is very weak. As in much behaviorist- inspired experimental research, the desire to have full control over the stimulus structure has led to the use of stimulus domains with low meaningfulness and hence poor ECOLOGICAL VALIDITY.

The third tradition derives from the application of psychological methods to LEXICAL SEMANTICS, the representation of word meaning. In this tradition, concepts are studied through their expression in commonly used words. Working within the Fregean tradition, interest has focussed on how the intensions of concepts are related to their extensions. Tasks have been devised to examine each of these two aspects of people's everyday concepts. Intensions are typically studied through feature listing tasks in which people are asked to list relevant aspects or attributes of a concept which might be involved in categorization, and then to judge their importance to the definition of the concept. Extensions are studied by asking people either to generate or to categorize lists of category members. The use of superordinate concepts (e.g. Birds or Tools) allows instances to be named with single words. Extensions may also be studied through the classification of hypothetical or counterfactual examples, or through using pictured objects.

Five broad classes of model have been proposed within this tradition. The classical model assumes that concepts are clearly defined by a conjunction of singly necessary and jointly sufficient attributes (Armstrong et al., 1983, Osherson and Smith, 1981). The first problem for this view is that the attributes which people list as true or relevant to a concept's definition frequently include non-necessary information which is not true of all category members (such as that birds can fly), and often fail to provide the basis of a necessary and sufficient classical definition. Second, there are category instances which show varying degrees of disagreement about their classification both between individuals, and for the same individuals on different occasions (McCloskey and Glucksberg, 1978). Third, clear category members differ in how typical they are judged to be of the category (Rosch, 1975). The classical view was therefore extended by proposing two kinds of attribute in concept representations -- defining features which form the core definition of the class, and characteristic features which are true of typical category members only and which may form the basis of a recognition procedure for quick categorization. Keil and

Batterman (1984) reported a development with age from the use of characteristic to defining features. The extended classical view however is still incompatible with the lack of clearly expressible definitions for most everyday concept terms.

The second model is the prototype model, (Rosch and Mervis, 1975). Concepts are represented by a prototype with all the most common attributes of the category, and instances belong in the category if they are sufficiently similar to this prototype. The typicality of an instance in a category depends on the number of attributes which an instance shares with other category members. Prototype representations lead naturally to non- defining attributes, and to the possibility of unstable categorization at the category borderline. Such effects have been demonstrated in a range of conceptual domains. A corollary of the prototype view is that the use of everyday concepts may show non-logical effects such as intransivity of categorization hierarchies, and non-intersective conjunctions (Hampton, 1982, 1988). Associated with prototype theory is the theory of basic levels in concept hierarchies. Rosch et al. (1976) proposed that the similarity structure of the world is such that we readily form a basic level of categorisation - typically that level corresponding to high frequency nouns like chair, lemon or car - and presented evidence that both adults and children find thinking to be easier at this level of generality (as opposed to superordinate levels such as furniture or fruit, or subordinate levels such as armchair or McIntosh apple.) This intuitively appealing notion has however proved hard to formalize in a rigorous way, and the evidence for basic levels outside the well-studied biological and artifact domains remains weak. Attempts to model the combination of prototype concept classes with FUZZY LOGIC (Zadeh, 1965) proved to be ill-founded (Osherson and Smith, 1981), but led to the development of more general research in CONCEPTUAL COMBINATION (Hampton, 1988; Rips, 1995).

The third model, the exemplar model, is only weakly represented in the lexical semantics research tradition. There have been proposals that lexical concepts could be based not on a prototype, but on a number of different exemplar representations. For example small metal spoons and large wooden spoons are considered more typical than small wooden spoons and large metal spoons (Medin and Shoben, 1988). This fact could be evidence for representation through stored exemplars, although it could also be explained by a disjunctive prototype representation. Formally explicit exemplar models are generally underpowered for representing lexical concepts, having no means to represent intensional information for stimulus domains that do not have a simple dimensional structure. As a result they have no way to derive logical entailments based on conceptual meaning (e.g. that all robins are birds).

The fourth model is the theory-based model (Murphy and Medin, 1985) which has strong connections with the cognitive development tradition. Concepts are embedded in theoretical understanding of the world. While a prototype representation of the concept BIRD would consist of a list of unconnected attributes, the theory-based representation would also represent theoretical knowledge about the relation of each attribute to others in a complex

network of causal and explanatory links, represented in a structured frame or schema. Birds have wings in order to fly, which allows them to nest in trees, which they do to escape predation, and so forth. According to this view, objects are categorized in the class which best explains the pattern of attributes which they possess (Rips, 1989).

The final model, psychological essentialism (Medin and Ortony, 1989), is a development of the classical and theory-based models, and attempts to align psychological models with the philosophical intuitions of Putnam and others. The model argues for a classical "core" definition for concepts, but one in which the core definition may frequently contain an empty "place holder". People believe that there is a real definition of what constitutes a bird (an essence of the category), but they don't know what it is. They are therefore forced to use available information to categorize the world, but remain willing to yield to more expert opinion. Psychological essentialism captures Putnam's intuition that people defer to experts when it comes to classifying biological or other technical kinds (e.g. gold). However it has not been shown that the model applies well to concepts beyond the range of biological and scientific terms (Kalish, 1995) or even to people's use of natural kind terms such as "water" (Malt, 1994).

The proliferation of different models for concept representation reflects both the diversity of research traditions, and the many different kinds of concept we possess and the different uses that we make of them.

-- James Hampton

REFERENCES

Armstrong, S.L., L. R. Gleitman, and H. Gleitman. (1983). What some concepts might not be. Cognition. 13: 263-308.

Barsalou, L.W. (1993). Structure, flexibility, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A.C.Collins, S.E.Gathercole, and M.A.Conway (Eds.) Theories of Memory. Hillsdale, NJ: Erlbaum.

Bruner, J.S., J. J. Goodnow, and G.A. Austin. (1956). A Study of Thinking. New York: Wiley.

Carey, S. (1985). Conceptual change in childhood. Cambridge, MA: MIT Press.

Fodor, J.A. (1983). The Modularity of Mind. Cambridge, MA: MIT Press.

Fodor, J.A. (1994). Concepts - a pot-boiler. Cognition. 50: 95-113.

Frege, G. (1952). On sense and reference. In P.Geach and M. Black (Eds.) Translations from the philosophical writings of Gottlob Frege. Oxford: Basil Blackwell.

Hampton, J.A. (1982). A Demonstration of Intransitivity in Natural Categories. Cognition. 12: 151-164.

Hampton, J.A. (1988). Overextension of conjunctive concepts: Evidence for a Unitary Model of Concept Typicality and Class Inclusion. Journal of Experimental Psychology: Learning, Memory and Cognition. 14: 12-32.

Kalish, C.W. (1995). Essentialism and graded membership in animal and artifact categories. Memory and Cognition. 23: 335-353.

Keil, F.C. (1989). Concepts, Kinds, and Cognitive Development Cambridge, MA: MIT Press.

Keil, F.C., and N. Batterman. (1984). A Characteristic-to-Defining Shift in the Development of Word Meaning. Journal of Verbal Learning and Verbal Behavior. 23: 221-236.

Kripke, S. (1972). Naming and necessity. In D. Davidson and G.Harman (Eds.) Semantics of Natural Language. Dordrecht: Reidel.

Malt, B.C. (1994). Water is not H2O. Cognitive Psychology. 27: 41-70.

McCloskey, M., and S. Glucksberg. (1978). Natural categories: Well-defined or fuzzy sets? Memory and Cognition. 6: 462-472.

Medin, D.L., and A. Ortony. (1989). Psychological essentialism. In S.Vosniadou and A.Ortony (Eds.) Similarity and analogical Reasoning Cambridge: Cambridge University Press, pp. 179-195.

Medin, D.L., and E. J. Shoben. (1988). Context and Structure in Conceptual Combination. Cognitive Psychology. 20: 158-190.

Murphy, G.L., and D. L. Medin. (1985). The role of theories in conceptual coherence. Psychological Review. 92: 289-316.

Nosofsky, R.M. (1988). Exemplar-based accounts of relations

between classification, recognition, and typicality. Journal of Experimental Psychology: Learning, Memory and Cognition. 14: 700-708.

Osherson, D.N., and E. E. Smith. (1981). On the adequacy of prototype theory as a theory of concepts. Cognition. 11: 35-58.

Piaget, J. (1967). Piaget's theory. In J.Mussen (Ed.) Carmichael's manual of child psychology (Vol 1.) New York: Basic Books.

Putnam, H. (1975). The meaning of 'meaning'. In his Mind, language and reality, vol. 2: Philosophical papers. Cambridge: Cambridge University Press.

Rips, L.J. (1989). Similarity, typicality and categorization. In S.Vosniadou and A. Ortony (Eds.) Similarity and analogical Reasoning Cambridge: Cambridge University Press, pp. 21-59.

Rips, L.J. (1995). The current status of research on concept combination. Mind and Language. 10: 72-104.

Rosch, E. (1975). Cognitive representations of semantic categories. Journal of Experimental Psychology: General. 104: 192-232.

Rosch, E., and C. B. Mervis. (1975). Family resemblances: studies in the internal structure of categories. Cognitive Psychology. 7: 573-605.

Zadeh, L. (1965). Fuzzy sets. Information and control. 8: 338-353.