© Blackwell Publishers Ltd. 1995, 108 Cowley Road, Oxford, OX4 1JF, UK
and 238 Main Street, Cambridge, MA 02142, USA.
Mind & Language, ISSN: 0268-1064
Vol. 10. No. 4 December 1995, pp 402–422.

Two Concepts of Concept

MUHAMMAD ALI KHALIDI

Abstract: Two main theories of concepts have emerged in the recent psychological literature: the Prototype Theory (which considers concepts to be self-contained lists of features) and the Theory Theory (which conceives of them as being embedded within larger theoretical networks). Experiments supporting the first theory usually differ substantially from those supporting the second, which suggests that these theories may be operating at different levels of explanation and dealing with different entities. A convergence is proposed between the Theory Theory and the intentional stance in the philosophy of language and mind. From this stance, concepts should not be thought of as concrete physical entities.

Philosophers discuss meaning, psychologists concepts. Psychologists experiment with subjects, philosophers speculate about agents. But underneath such terminological and methodological differences lie some common concerns. In recent psychological work, there are a number of important results relevant to philosophical concerns about concepts and lexical meaning. I will be focusing on a recent debate in the psychological literature and will argue that it can be illuminated by a philosophical account of concepts. For almost a decade now, the dominant accounts of concepts in cognitive psychology—Prototype Theory and its close relations—have been challenged by an alternative, which has been called the 'Theory Theory' of concepts. The central difference between the two views is that the Prototype Theory conceives of concepts as lists of features or attributes, while the Theory Theory thinks of them as being enmeshed in a more comprehensive theoretical network.[1] But

I am grateful to Akeel Bilgrami, Jesse Prinz, Terry Regier, Richard Rosenblatt, Josef Stern, and two anonymous referees for very helpful comments on earlier versions. Thanks also to audiences at the University of Chicago and the University of California at Berkeley for feedback. Some of these ideas were hatched during a N.E.H. Summer Seminar on Mental Representation directed by Rob Cummins, with additional (patient) instruction in cognitive psychology by Denise Cummins.
Address for correspondence: Department of Philosophy, 1050 East 59th Street, University of Chicago, Chicago, IL 60637, USA.
Email: ma-khalidi@uchicago.edu.

[1] The Prototype Theory means different things to different people. In my characterization, I am following one fairly prevalent construal of the Prototype Theory; other authors have interpreted it as an exemplar-based rather than a feature-based model. But even exemplars are reduced to clusters of features on some accounts, since feature-matching seems to be the least controversial way of implementing the categorization model of the Prototype Theory, see e.g. Smith and Medin, 1981, pp. 149–50.

this way of putting things makes the differences sound more superficial than they really are. It may be more accurate to say that the two theories view concepts from divergent perspectives and may therefore be talking about different things. These perspectives are what Dennett has called the *design stance* and the *intentional stance,* respectively. After proposing this way of construing the difference between these two psychological theories, I will argue for a convergence between an intentional, holistic account of concepts and the account introduced by the Theory Theory. This philosophically inspired account of concepts can be used to show how concepts can be given determinate identity conditions, thus answering an objection regarding the Theory Theory's manner of individuating concepts.

1. The Prototype Theory and the Theory Theory

The Prototype Theory emerged partly as a result of findings of typicality in the responses of subjects while performing certain cognitive tasks. These experimental results confirmed an intuitively appealing claim which Rosch and Mervis put as follows (1975, p. 573): 'As speakers of our language and members of our culture, we know that a chair is a more reasonable exemplar of the category *furniture* than a radio, and that some chairs fit our idea or image of a chair better than others.' But 'typicality effects' in cognition are not confined to explicit avowals by subjects as to which exemplars they consider more typical of a certain concept. The instances judged more typical by participants in psychological experiments turn out to be implicated in results which involve cognitive effects of a 'deeper' nature.

There are three main types of experiments which have been taken as evidence for typicality effects. In the first, a group of subjects is asked to rate the extent to which an instance represents their 'idea or image of the meaning of the category name' (ibid, p. 588). For instance, they are given such words as 'robin', 'bluebird', 'seagull', 'penguin', and 'chicken', and they are asked to rate their typicality as instances of the concept *bird* on a scale from 1 to 7. In one of Rosch's experiments, subjects were instructed that 'some reds are redder than others' and that a Pekinese is a 'less doggy dog' than a Retriever or a German Shepherd. The instructions also read, in part: 'Don't worry about *why* you feel that something is or isn't a good example of the category ... Just mark it the way you see it.' (Ibid., p. 589, emphasis added) The typicality judgements of these subjects were averaged out and tabulated. In another experiment, subjects were supplied with the name of a particular kind of bird with the instruction to list as many properties of that bird as possible in a short time period, say 90 seconds; for example 'has feathers', 'flies', 'sings' and so on. The birds already judged more typical by the first set of subjects shared more features with other birds than the ones judged less typical. In Rosch's terminology, they had a higher degree of *family resemblance.* Finally, in another kind of experiment, Rosch's subjects were asked to respond true or false to such statements as 'A robin is a bird', 'A penguin

© *Blackwell Publishers Ltd. 1995*

is a bird' and 'A lion is a bird', and their reaction times and error rates are measured (Rosch, 1978, p. 38). Faster reaction times were recorded for sentences involving the instances judged more typical in the first experiment. In addition, Rosch claims that the words for typical instances are likely to be named first and more frequently when subjects are asked to list instances of a certain concept, and the words for typical instances are the first ones to be learned by children and are learned more quickly by them (ibid., pp. 38–9).[2]

These experiments show a strong correlation between three different measures: typicality judgments, degrees of family resemblance, and reaction times for categorization decisions and similar cognitive tasks. Typicality effects do not just concern subjects' implicit or explicit beliefs about which instances are more culturally salient; they are supposed to tell us something more significant about human conceptual organization or about the nature of concepts themselves. Accordingly, it was postulated that many of our concepts consist in prototypes, weighted clusters of features characteristic of each concept. Rather than a list of necessary and jointly sufficient features, a prototype is considered to be a probabilistic feature list, with each feature being weighted according to its importance to that concept. The more features an instance shares with the concept and the more important those features, the more prototypical that instance is of the relevant concept.

But the Prototype Theory has recently come in for some criticism. In a backlash against this once-dominant account of concepts, a number of cognitive psychologists have begun to argue that psychological concepts are more enmeshed in relevant theories and couched in explanatory beliefs. These advocates of the Theory Theory of concepts do not regard concepts as being relatively independent, self-contained entities in the manner of the Prototype Theory. The move to theory was spurred by a number of phenomena which are hard to account for on psychological models that treat concepts as collections of features, even probabilistic collections of features. There are two main cognitive effects that do not comport well with such models. The first is that the kinds of features that subjects associate with certain concepts vary widely and almost without limit when one varies the experimental context in which they are tested. Rather than accessing a fixed set of features in conjunction with each concept, there is apparently no limit to the features that even a single subject associates with a certain concept depending on the context in question. Barsalou (1982) found that common features in similarity judgments can appear and disappear with context. Similarly, a well-known result due to Barclay et al. (1974), found that different features were associ-

[2] A referee points out that words that are named first and learned first by children are not necessarily the most typical, since kindergartners are more likely to know the name for penguin than that for sparrow. This suggestion has some plausibility, but it goes against the results cited in Rosch, 1978. An experimental adjudication of the issue seems to be in order.

© *Blackwell Publishers Ltd. 1995*

Furniture		**Bird**	
weight	*feature*	*weight*	*feature*
1.0	physical object	1.0	moves
1.0	non-living	1.0	has wings
0.9	decorative	1.0	has feathers
0.8	rigid	0.8	flies
0.7	has legs	0.6	sings
0.5	has seat	0.5	small size
	etc.		etc.

General Processing Assumption: An entity X is categorized as an instance or subset of concept Y if and only if X possesses some critical sum of the weighted features of Y.

Prototypicality effects arise because an instance that has more of the important features is judged by the subject to have the critical sum more quickly, so it is categorized more rapidly. For instance, a robin presumably achieves the critical sum for the concept *bird* more quickly than a penguin.

Figure 1 Adapted from Smith and Medin (1981), taking Prototype Theory as the featural version of the Probabilistic View of concepts. (Note that Smith and Medin consider the Prototype Theory as being a version of what they call the 'Exemplar View', a minority interpretation among psychologists, most of whom consider it a version of what they call the 'Probabilistic View'.)

ated with the concept *piano* in the context of producing music and that of moving furniture.

A second difficulty for the Prototype Theory is the ability of subjects to make cross-conceptual links and to relate their beliefs involving different concepts in informative ways, abilities that are not easily explained on a model of concepts as bounded, self-contained feature lists. Categorization is not a simple matter of matching features among a concept and its instances, but is determined by inferential processes driven by surrounding explanatory theories. Instances that contain correlated rather than unrelated features are categorized more efficiently and variations in the context can lead to different grounds for classifying instances. Following Barsalou (1993), these two related features of concepts can be dubbed, *flexibility* and *structure*, respectively.

It is significant that the experiments taken to support this new approach to concepts are rather different in character from those that provided evidence for the Prototype Theory. The typicality effects I mentioned emerge most clearly under time pressure and in tasks involving routine categorization decisions and identification of instances, when subjects are not questioned as to the reasons behind their decisions.[3] In the psychological exper-

[3] Not all such experiments involve measuring reaction time (RT), but when RT is not measured, there is usually either time pressure, or else the tasks consist of routine categorization judgements or rote memorization. An example of an experiment without RT

© *Blackwell Publishers Ltd. 1995*

iments that support the Theory Theory, by contrast, subjects are typically presented with full-blown narratives or accounts of natural processes and then asked various questions about them. The data in these cases consist of what psychologists call 'protocol analyses': verbatim transcripts of subjects' responses and their attempts to justify those responses under the scrutiny of an experimenter. Neither the categorization tasks nor the subsequent justifications are subject to time constraints, and the categorizations are seldom as routine as those that occur in the experiments just described.

In Keil's work on conceptual development in children, the methodology is more Piagetian in character, since a prominent role is given to extensive interviews with subjects. A story or narrative consisting of several sentences is read out to a child who is then asked to categorize something. Depending on the child's answer to this question, the experimenter goes on to ask a number of follow-up questions in order to elicit the child's rationale for the classification in question. The experiments reported in Keil (1989b) were designed to determine whether certain concepts are constituted by simple feature lists or by more global theories. In one experiment, children are read a story about animals living on a farm. They are told that the animals neigh, eat oats and hay, and that people saddle them and ride them, but they are also told that they were examined by scientists and found to have the insides of cows, the blood and bones of cows, and that their parents and offspring were found to be cows. The children are asked what they think these animals really are, horses or cows. They are encouraged to justify their judgments during extensive conversations, in an effort to examine their corresponding concepts. In another experiment, Keil tells a story about taking a raccoon, shaving away some of its fur, dyeing it black with a single white stripe down the center of its back, then inserting a sac of smelly odor into its body. The child is asked whether the resulting animal is a raccoon or skunk.

Keil finds that for many ordinary natural kind and artifact concepts, there is a significant shift from reliance on superficial features to one on deeper explanatory features, which occurs at different ages for different concepts, as early as the preschool years for some concepts, and as late as the fourth

---

measurements, but where time pressure is clearly a factor is McCloskey & Glucksberg, 1978. Subjects were asked to make yes–no category membership judgments for each of 540 exemplar-category name pairs; most subjects completed the task within 50 minutes. They were instructed 'to take enough time for each pair, but not to linger over any individual item' (ibid., p. 464). An example of a rote memorization task is Keller & Kellas, 1978, in which it was found that more typical items were recalled better than atypical items from word lists that subjects were exposed to. A possible exception to this pattern is Rips, 1975, which found that typicality influenced the inductive generalizations that subjects were willing to make (there was no time pressure and judgments were not routine). But in later work, Rips has argued that subjects' similarity judgments do not correlate well with their categorization judgements, see e.g. Rips & Collins, 1993. However, it should be emphasized that the issue of similarity-based and rule-based categorization is not identical with prototype-based and theory-based categorization, since Prototype Theory can be made compatible with both similarity- and rule-based models.

© *Blackwell Publishers Ltd. 1995*

grade for others.[4] After the shift, subjects deploy sophisticated causal theories in performing categorization tasks, rather than mere characteristic features or prototypes. While he allows that characteristic feature lists and prototypes are 'certainly associated with how we often use concepts and normally rapidly identify their instances', the interviews he conducts reveal an acquired reliance on more sophisticated theoretical frameworks (Keil, 1986, p. 152). Given suitably bizarre contexts, children of a certain age will rely on theories of reproductive descent, internal structure, and other such explanatory frameworks to categorize animals, rather than simply appealing to such superficial characteristics as color, shape, and outward appearance. Similar evidence of structure is found for other kinds of concepts, for example, artifacts.

These experimental results tally better with a picture according to which concepts are embedded in a total framework of explanatory beliefs (or theories), which one draws upon in part in performing a particular cognitive task—with different parts of the entire corpus invoked in different tasks, even ones involving a single concept. As Murphy and Medin put it in a seminal article (1985, pp. 289–92): '[C]urrent ideas, maxims, and theories concerning the structure of concepts ... are inadequate, in part, because they fail to represent intra- and inter-concept relations and more general world knowledge. We propose a different approach in which attention is focused on people's theories about the world ... [We] wish to reduce the importance of individual attributes in conceptual representations and to emphasize the interaction of concepts in theory-like mental structures.' Similarly, Neisser (1987, p. 9) comments: 'Since Rosch's original discoveries, the field as a whole seems to be moving from an emphasis on objective attributes and similarity to a more recent insistence on the role of theories and idealized models.' Although a full-blown Theory Theory has yet to emerge, there is dissatisfaction with a view of concepts as self-contained psychological structures, relatively isolated from one another, and from pertinent background beliefs. Generally speaking, cognitive tasks that involve explaining and justifying classifications in a specific context, rather than rapid categorization decisions without such a context, have required psychologists to posit an interrelated network of conceptual information rather than independent collections of feature lists. The less routine the categorization task, the more it seems as though concepts are embedded in larger theoretical networks with a dense pattern of correlations linking one concept to another.

---

[4] He usually talks about a shift from 'characteristic features' to 'defining features', but he also explains: 'The distinction can be recast, however, in a form more compatible with [Quine's outlook]: it can be viewed as a shift from general atheoretical relations motivated by content-independent principles of similarity based on simple perceptual comparisons and typicality calculations, to theoretically organized relations that for these special (nominal kind) terms appear to yield defining features.' (Ibid., p. 269)

© *Blackwell Publishers Ltd. 1995*

2. *Comparison and Evaluation*

The attempt to compare these two theories of concepts raises an interesting methodological question as to the relation between them. One could, of course, treat the two theories as rivals providing competing accounts of concepts—and that is how they are normally regarded. But there are other ways of construing the relation between them which do not consider them to be competitors; two in particular may be worth exploring.

One way of reconciling the two theories is to regard the Prototype Theory as measuring cognitive effects that emerge in certain special cases when we are deploying a particular kind of default theory. I have already said that typicality effects are not only a matter of subjects' responses to questions about whether, say, a Retriever or a Pekinese were a more typical dog. If that were all there was to typicality effects, then these beliefs could be explained without the invocation of prototypes. One would say that members of a culture have certain implicit or explicit beliefs about which instances of a concept are more salient, are of a common-or-garden variety, are more often used in instructing children, appear in illustrations alongside dictionary definitions, and so on. However, experimental results such as those involving ease and rapidity of categorization may not seem to be easily explained on the theoretical approach to concepts, since they suggest that some concepts are more cognitively accessible than others and some instances are more readily categorizable. But this may be accounted for by noting that some conceptual information is more diagnostically useful because it is associated with our common social, cultural, and geographic settings. In light of this, we might evolve certain cognitive short-cuts to access some chunks of our entire corpus of beliefs rather than others in certain contexts. There may be certain background theories that we fall back on in certain settings and they may be the ones that have served us particularly well in our most habitual surroundings. These could be the theories we rely on in our initial reactions to a perceptual stimulus and those we resort to under time pressure when no context has been specified in a laboratory setting. The Theory Theory does not provide us with a ready-made way of explaining reaction times and similar psychological phenomena, but it may be developed to deal with these results if prototype effects are thought of as arising because of certain default theories.

Interestingly, Rosch herself hints at something like this interpretation of the Prototype Theory in one of her original papers. She points out that there will surely be context effects determining which items are named, listed, or expected when subjects are given the names of categories. In response to the objection that prototype findings are only relevant to the artificial situation of the laboratory in which no context is specified, she states that her findings reveal something about the context that the subjects themselves contribute. Rosch points out (1978, p. 43): 'in the absence of a specified context, subjects assume what they consider the normal context or situation for occurrence of that object.' While she acknowledges that the effects that she has measured

© *Blackwell Publishers Ltd. 1995*

may be radically context-dependent, she does not consider what would happen if one were to specify different contexts.

However, viewing prototypes as default theories involves changing our whole conception of what prototypes are; it effectively denies the existence of such entities as prototypes, preserving only prototypical *effects*. Therefore, I will propose another way of relating the two theories, by taking them to be theories of different aspects of our cognitive abilities. The fact that the kinds of experiments that are taken to supply evidence for the two theories are so different might suggest that they are tracking different kinds of phenomena. The theoretical account of concepts would seem more suitable for discussing agents' deliberative decision-making procedures, inferential reasoning involving articulated bodies of information, and the comprehension and production of narratives, but less relevant to other tasks. The latter might include reactions to words or images after brief exposure, automatic perceptual judgments regarding the environment, and the spontaneous generation of lists of words in response to certain brief questions.

There is some evidence of such a distinction in the psychological literature. In a related context, Medin and Wattenmaker (1987) invoke a distinction between 'basic learning, memorial, and perceptual processes' on the one hand and 'higher cognitive processes such as theory building' on the other (p. 55). A similar distinction is hinted at in a paper by Murphy and Medin (1985), in which they acknowledge that little theoretical knowledge is probably invoked when we classify something as a robin, whereas rather more is required with novel objects and borderline cases, and when the categorization must be justified and explained (p. 296). This could also be understood using a distinction often made by psychologists between a concept's *core* and its *identification procedures*. The core of a concept may be theoretical, whereas the procedure for identifying a given instance of a concept may be prototypical. Gleitman, Armstrong and Gleitman (1983) speculate that the graded judgments are a function of a mentally stored identification procedure used to sort through things quickly, whereas the core is what determines membership in a category (i.e. upon reflection).[5] The difference between the first process and the second is that the first conceives of the subject as an automatic detector or categorizer of the environment, whereas the second takes the subject as a rational agent who formulates theories about the environment and responds to it through the filter of those theories. The systems tracked by each account of concepts may constitute two different aspects of the human cognizer and may correspond to two different ways of theorizing about human cognition. They may be descriptions of our cognitive abilities at different levels of explanation.

---

[5] But note that in another paper, Armstrong, Gleitman and Gleitman (1983) are ultimately unhappy with this reconciliation, saying that identification procedures are likely to involve something other than prototypes, since identifying instances of a concept is not merely a matter of consulting lists of perceptual features. (p. 298)

© *Blackwell Publishers Ltd. 1995*

Hence, there are at least two ways of construing the relationship between the Theory Theory and the Prototype Theory, which do not involve considering them as competitive rivals. In the first case, the Prototype Theory would just be a special case of the Theory Theory. In the second case, the Prototype Theory and the Theory Theory take entirely different stances towards psychological subjects and may be isolating different entities. On both these interpretations, I would argue that the Theory Theory's account of concepts is given from what Dennett has termed the 'intentional stance'. Thus, it is important to make clear what it means to take the intentional stance towards the mind; in the following section, I will outline an account of concepts as seen from the intentional stance.

3. *Concepts from the Intentional Stance*

The distinction between the design stance and the intentional stance towards the study of the mind is characterized by Dennett as follows. From the design stance, 'one ignores the actual (possibly messy) physical constitution of an object, and, on the assumption that it has a certain design, predicts that it will behave *as it is designed to behave* under various circumstances.' (1987, pp. 16–17) By contrast, on the intentional stance, 'first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose.' (Ibid., p. 17) How does this distinction help to explain the difference between the Prototype Theory of concepts and the Theory Theory? According to the second proposal I made in the previous section, the former theory thinks of concepts as being constituted from a bundle of features, and it thinks of concepts as being manifested in the organism when those features are detected in the world. The organism is designed in such a way that whenever a certain number of those features is detected and a critical sum is attained, the corresponding concept is tokened. Moreover, the features may even be considered to be perceptual ones.[6] On the theoretical view of concepts, by contrast, the organism is regarded as an agent which has formed rational beliefs about the environment and reasons about the world in conformity with those beliefs. Concepts, from this perspective, are simply components of fully-fledged beliefs that have been ascribed to subjects according to our usual ascriptive practices.

Not only are the concepts discussed by the Theory Theory entities posited from the perspective of the intentional stance, I would argue that they con-

[6] Armstrong, Gleitman, and Gleitman observe (1983, p. 271): 'To the extent that the prototype views are still componential, they still give hope of limiting the primitive basis, the set of innate concepts. If correct, they allow the empiricist program to go through in detail for the complicated concepts.'

© *Blackwell Publishers Ltd. 1995*

form broadly to a holistic account of belief and meaning. The main features of the Theory Theory are reminiscent of a Quinean view in the philosophy of language, but without sharing Quine's suspicion of intentional entities. This is not entirely a coincidence since at least some of the psychologists mentioned have been explicitly influenced by Quine's work, as witnessed by Keil's use of the phrase 'the web of belief'. In this section I will draw out some of the points of convergence between a holistic account of concepts and that provided by the Theory Theory, in the process further justifying the conjecture that it is embedded in the intentional rather than the design stance. This attempt is also important because it will help us (in the following section) to resolve a major problem for the Theory Theory's account of concepts, namely one concerning the source of stability for concepts and individuation conditions for them.

Although the original source of the intentional stance may be found in the work of Quine and Davidson, neither of these philosophers is prone to talk in terms of concepts. So, I will try to elaborate an account of concepts that is self-standing, though it owes its inspiration to their methodological framework. That framework locates the notions of meaning and belief in the process of translation or interpretation.[7] The paradigm case of the interpretive process for Quine is, of course, the encounter between the linguist and the informant in the field. By observing the informant's utterances and actions, the linguist attempts to frame 'analytic hypotheses' about the mean-

[7] For Quine's view, see Quine, 1960, and for Davidson's many of the papers collected in Davidson, 1984. Some of their assumptions and attitudes are shared by a number of other philosophers, notably, Lewis, 1974, Haugeland, 1978, Dennett, 1987, and Bilgrami, 1992. A number of important philosophical issues will be bracketed for these purposes; two are worth mentioning to be on the safe side. First, a question arises as to what such a translation or interpretation is concerned to preserve. The answer cannot be truth, since there are points at which the agent being interpreted is likely to come out false by the lights of the agent doing the interpreting. A more promising candidate is rationality, though it is harder to say exactly how this is to be understood and where to draw the limits of the rational. There are several attempts in the literature to specify what the interpretive process should capture and to encapsulate it in one or more interpretive principles. The Principle of Charity, the Principle of Humanity, and our 'general theory of persons', are just three of these attempts (drawn, respectively, from Davidson, 1984, Grandy, 1973, and Lewis, 1974). For these purposes, I will just assume that we have a fairly clear idea of what the translation aims to preserve and that we can construct one that does the job. This leads to the second issue, which concerns the question of indeterminacy. What ensures that there will be only *one* translation function that can deliver the goods? I will be making the assumption that the problem of indeterminacy is merely that of the underdetermination of theory by evidence, as applied to semantics. Just as there are constraints on our scientific theories which enable us to rule out what appear to be empirically adequate alternatives, there will be constraints which serve to show that one translation is superior to the others and enable us to rule that it is optimal. On this score, Lewis, 1974, Putnam, 1975, and Chomsky, 1980, are all illuminating. In short, unlike Quine, I will take it that the phenomenon of indeterminacy does not have debilitating consequences for semantics or psychology.

© *Blackwell Publishers Ltd. 1995*

ings of the informant's terms. These hypotheses will, for example, link up the informant's term 'gavagai' with the linguist's term 'rabbit'. After testing these hypotheses in the face of the empirical evidence, the linguist emerges with a complete translation manual which connects the informant's terms to the linguist's terms. For the purpose at hand, I will be stressing two aspects of the intentional stance: the inextricability of meaning and belief, and semantic holism. A number of important philosophical issues will be bracketed in the interest of focusing on the subject of concepts.

The inextricability of meaning and belief is critical for this account of concepts. Suppose we are confronted with an agent who utters something which we translate as 'Rabbits are insects'. If we are reasonably sure about our translations of plurals and the third person singular form of the verb 'to be', we still have (at least) two degrees of freedom in interpreting the sentence. We might decide to attribute the *false* belief that rabbits are insects, standing by our translations of the informant's words for rabbit and insect. Alternatively, we might revise one of our initial conjectures, perhaps ruling that 'gavagai' means not rabbit but rabbit-fly, since rabbits in this locale are fly-infested and previous utterances by our informant could have been alerting us to the presence of rabbit-flies in the vicinity rather than rabbits. That would enable us to attribute the *true* belief that rabbit-flies are insects. More evidence is needed to decide between the two courses. This illustrates the point that we have no handle on meaning which is independent of our handle on belief in this theoretical framework. The same body of evidence and the same process that enable us to attribute one also enable us to attribute the other.

As for semantic holism, the clearest way of casting that doctrine is as a denial of *atomism*: the idea that terms have the meanings that they do by virtue of *direct* relations to extra-linguistic determinants. These candidates for extra-linguistic determinants might be abstract entities, or they might be objects in the world external to the agent, or they might be well-defined structures in the agent's brain. Rather, according to interpretivism, terms have the meanings that they do because they are used in certain ways by the agents involved and they play a certain role in the agent's psychological economy. The relation of this holistic claim to the ones vetted above is fairly clear. It is on the basis of utterances and actions that we attribute certain meanings and beliefs to our informants, and this is what grounds the notions of meaning and belief. In the hypothetical case of the rabbit versus the rabbit-fly we will make the judgement in favour of one and against the other, not on the basis of some single relation that the agent may have (or may have had) to the external world, or even any single piece of evidence, but rather, the totality of the evidence. We aim at constructing a mapping between our respective vocabularies that exhibits a certain overall fit. This overall fit concerns a term's position in the linguistic practice of the agent being interpreted, in short, its place in that person's entire corpus of beliefs and intentional actions.

Where do concepts appear, if at all, on this picture? Linguistic or lexical

© *Blackwell Publishers Ltd. 1995*

concepts are interchangeable with meanings,[8] when we allow for the syntactic and stylistic infelicities that might result from substituting 'concept' for 'meaning' throughout this account. To say that 'gavagai' means rabbit is to say that our informant has the concept *rabbit* (or, as we sometimes say, has the concept *of a rabbit*). And we judge that our informant has the concept *rabbit* when and only when we have translated one of our informant's terms by our term 'rabbit' and interpreted some utterances as rabbit-utterances. Once we have done this, we have, willy-nilly, attributed the concept *rabbit* to our informant.

It might be asked how this account of concepts differs from the old cluster theory of concepts or of meanings, according to which every concept is a cluster of singly necessary and jointly sufficient features or attributes. In a more sophisticated version, one can work into it certain probabilistic measures to suggest that all attributes are not as important to the concept, a view very much in line with the Prototype Theory of concepts. There is no denying that for each linguistic concept there will be a cluster of shared beliefs held in common between interpreter and interpretee, for example all those rabbit-beliefs which they happen to share. One might even say that we attributed the concept *rabbit* on the basis of those beliefs, but that does not mean that it would always be the same set of beliefs; for someone else with a different total theory, we might attribute the same concept, even though we share a different subset of beliefs. The interpretation is not being driven by the presence of a requisite set of beliefs or features but by the need to make overall sense of the informant in an intentionalistic idiom. In other words, the ascription of concepts is subordinated to the need to make sense of the rational agent; the agent is not viewed merely as a complex feature detector, as on the design stance.

There is another crucial difference between this theory and a weighted cluster theory: there is no fixed cluster associated with each concept, since the concept inheres in the theoretical framework as a whole. We could isolate all the beliefs in which the associated term appears and regard this as the cluster, but there will be a *potential* infinitude of such beliefs, and each one of them will point to yet other beliefs, facts which make the cluster metaphor misleading. Notice that this corresponds to the feature of concepts that psychologists have dubbed their 'flexibility' or 'open-endedness', since there is no fixed set of features associated with each concept. On this view, the exist-

[8] This identification of lexical concepts with meanings is by no means unique to the interpretivist view, nor does it seem so controversial. It is also shared by many cognitive psychologists. For example, Carey writes (1988, p. 167n): 'I will use "concept x" and "meaning of the term x" interchangeably.' She goes on to say that in previous work, 'In every case that I found a difference in meaning of a term "x" between the child's lexicon and the adult's, there was a corresponding difference in the concept x, as revealed by patterns of inductive projection, sorting tasks, and other tasks not requiring the use of the term.' Similarly, Gleitman, Armstrong, and Gleitman state (1983, p. 88): 'for present purposes we make no fine distinction between theories of word meaning and theories of concept structure.'

© *Blackwell Publishers Ltd. 1995*

ence of concepts is bound up with the fact that beliefs are linguistically ascribed, for a concept marks a certain semantic feature common to all the beliefs in which it occurs. For example, our concept *rabbit* emerges from all the (potentially infinite) beliefs we have about rabbits. The link between concepts and terms may not be absolutely tight, for some linguistic concepts may not be expressible by single terms and some single terms may be conceptually polysemous, but there is no denying that on this account, concepts are ascribed on the basis of the way that linguistic terms are used. This clearly suggests that we are discussing *lexical* concepts, but that is also what cognitive psychologists are discussing in the experiments I have described.

It may be said that the difference between the intentional stance and the approach of the Theory Theory is that the former concentrates on the characterization of a whole mental life, whereas the latter is only interested in ascribing a handful of concepts in any one experiment. In the experiments that support the Theory Theory, psychologists attribute concepts singly and piecemeal based on specific questions that their subject are asked to answer under controlled circumstances, not based on some overall interpretive characterization of their subjects' psychological states. But just because psychologists concentrate at any one time on one or a small number of concepts, that does not mean that a larger mental life is not presumed to be in place, for implicit assumptions are also being made about how subjects are using terms and manipulating concepts that are *not* under direct investigation. Each of our concepts has more or less tenuous connections to other parts of our conceptual repertoire, and these connections can be made explicit given the right line of questioning. Indeed, this would seem to be one of the main effects of supplying subjects with bizarre contexts and requiring them to make classification judgements in unfamiliar settings. Even though a complete interpretation is not usually specified, that does not mean that such an interpretation is not presupposed in the background.

4. Holism and the Individuation of Concepts

The next step is to consider how the interpretive account of concepts might help us to resolve the 'circularity problem' for the Theory Theory: the problem of individuating concepts, providing them with some stability in spite of their highly flexible nature. I have been emphasizing the open-ended character of concepts on the Theory Theory and have contrasted it with the bounded, self-contained feature lists posited by the Prototype Theory. But the very unboundedness that furnishes much of the appeal of the Theory Theory also creates a problem for the individuation conditions of concepts. If, as Keil puts it (1989a, p. 49), 'concepts may only be understood in terms of the theories they are embedded in and theories only in terms of the concepts they embed', how are concepts to be disentangled from theories and how do they acquire stability and individuation conditions? On the Prototype Theory, it was fairly easy to individuate concepts and to provide them

© *Blackwell Publishers Ltd. 1995*

with determinate identity conditions. That was precisely because they were well-defined, bounded structures that were supposed to remain fairly constant across different contexts. But now that this picture has been found wanting for many cognitive tasks and it has been replaced with more flexibility and less stability, we no longer have a ready way of individuating concepts. Proponents of the Theory Theory have recognized this difficulty, but have not yet proposed a satisfactory answer.

This problem is closely related to what Murphy and Medin label the 'circularity objection', summarizing it very succinctly as follows (1985, p. 313): 'How can mental theories explain concepts . . . when theories themselves are made out of concepts?' In response, they write (ibid., p. 313): 'Concepts and theories must live in harmony in the same mental space; they therefore constrain each other both in content and in representational format.' Similarly, Keil responds to this difficulty with a metaphor: concepts are 'spiders in the web of belief' (1989a, p. 49). However, these brief characterizations do not supply a satisfactory answer to the problem. Fodor (1994) has focused on this very difficulty with the Theory Theory. He claims that the Theory Theory 'says that you have an essentially different *concept* of electrons from mine if . . . you have an essentially different *theory* of electrons from mine', and surmises that the problem of how to individuate concepts reduces to the problem of how to individuate theories, but he goes on to claim that 'nobody knows how to individuate theories' (pp. 110–11). As if to vindicate Fodor's claim, more than one advocate of the Theory Theory has been led to a conclusion of incommensurability about concepts associated with different theories or successive developmental stages in children. Gopnik (1988) has said that children's concepts of object-permanence, space, and object-identity 'are inextricably intertwined with other concepts in the theory', adding that 'all of them will change as the theory changes' (p. 205). Carey (1988) also holds that children's concepts are 'locally incommensurable' with those of adults, though she thinks that this does not preclude communication between adults and children. Thus, there seems to be a danger on the Theory Theory that concepts will be pictured to be so closely intertwined with theories that they will no longer be separable from them at all and cannot be independently individuated. Can the intentional stance help to solve this problem?

Rather than likening concepts to 'spiders in the web of belief', as Keil does, an alternative account can be given by taking more seriously an analogy with economics. One often speaks of an agent's 'mental economy' and of a concept as having a certain value in that mental economy. I propose to take this metaphor to heart by comparing a theory to an economic system and a concept to the value of the currency employed within that system.[9] On the interpretive view, concepts or meanings feature in comparisons undertaken

[9] The analogy between economic value and semantic value has also been discussed briefly by Dennett (1987, p. 208): he compares the problem of ascribing beliefs and concepts to different agents to that of figuring the values of commodities in different currencies.

© *Blackwell Publishers Ltd. 1995*

between two agents, just as currency values are needed primarily to determine exchange rates. The value of a certain currency emerges from the structure of the economy in which that currency is employed and the way to determine it is by comparing one economy with another, that is, by determining the exchange rate between the two economies. A radical interpreter can be likened to a tourist in a foreign country who knows neither the exchange rate nor the local market value of commodities and must discover both at once on the basis of the available evidence. As a first rough guess, the tourist might assume that things have the same value they have back home (which is analogous to assuming agreement on beliefs). But in a land where we know that coffee is scarce, we won't assume that the price of a cup of coffee is the same as it is in our country. We might begin by estimating that it costs twice as much and set a tentative exchange rate based on this estimate. After we have provisionally fixed the exchange rate, we may still expect a sandwich to cost roughly as much as it does at home. If it does not, our estimate for the exchange rate may have to be revised. The task of figuring a 'fair' exchange rate between two economies is analogous to the problem of comparing two agents' sets of concepts. There are also certain disanalogies, of course. For one thing, in an economic system, there is only one exchange rate to determine and, once discovered, it gives us a complete 'translation' between the two systems. In the case of belief systems, there are numerous concepts to be matched up and each match constrains, but does not force, other matches.

The concept-sharing relation is a derivative one on this account. If two agents share the concept *rabbit*, that does not tell us anything specific about the beliefs that they share or any particular additional fact about them. Instead, it marks a certain similarity in their mental economies. But there is nothing vague or unstable about this relation: agents have the concept *rabbit* by virtue of the fact that we use our term 'rabbit' in translating their utterances or ascribing their beliefs. This gives us a fully determinate way of saying whether an individual does or does not possess a certain concept. Such an interpretive decision is not made on the grounds that the term features in all the same beliefs, or even a specific subset of beliefs, but it does emerge out of a definite process of interpretation. This shows that concepts are not mere sets of beliefs or features and captures their open-ended character, while at the same time giving us a determinate method for ruling whether a certain concept is present or not. Individual interpretive decisions may not be easy, as can be seen from some of Keil's experimental protocols, in which the experimenter is deciding whether to ascribe such concepts as *racoon* or *skunk* to small children (see Figure 2). Still, the interpretive process and the standards applied by psychologists and other interpreters give us a way of picking out concepts which does not *reduce* them to specific theories or sets of beliefs. The translation function between the interpreter and interpretee is what provides concepts with stability and allows us to disentangle them from theories.

At this point it is worth raising another objection to the account of con-

© *Blackwell Publishers Ltd. 1995*

C: They made it into a skunk.
E: Why do you think it's a skunk?
C: Because I just do.
E: Why do you think it's a skunk?
C: Because they MADE IT INTO A SKUNK.
E: What about it makes it a skunk?
C: I don't know.
E: Could it still be a raccoon or did they make it into a skunk?
C: Inside it's, I guess it's a . . . a. I guess it's a raccoon.
E: Well, which do you think the animal really is? Do you think it's really a raccoon or do you think it's really a skunk?
C: A raccoon!
E: Can it be a . . .
C: (interrupting) It's a skunk.
E: Which do you really mean?
C: A skunk.
E: Can it be a skunk if its mommies and daddies were raccoons?
C: Yes.
E: Can it be a skunk if its babies were raccoons?
C: Yes.
E: (repeats entire story) Which do you think it really was?
C: A skunk. Because it looks like a skunk, it smells like a skunk, it acts like a skunk, and it sounds like a skunk. (The child was not told this).
E: So it can be a skunk even though its babies are raccoons?
C: Yes!

Figure 2 Example of an experimental protocol taken from Keil (1989, p. 188). This conversation between the kindergarten child (C) and the experimenter (E) was conducted after the raccoon/skunk story was read to the child.

cepts that emerges from the intentional stance. It may be said that the theory-embeddedness of concepts has the consequence that concepts are peculiar entities with properties that we would not associate with common-or-garden variety concrete objects. This challenge to the interpretivist view might be made more precise by focusing on its holistic account of concepts. The case against holism has been made forcefully in a recent work by Fodor and Lepore (1991). The upshot of that argument is that the doctrine of holism appears to rule out a notion of identity of meaning across different believers or in the same believer at different times. Briefly, the reasoning is as follows. Holism states that the meaning of any term in an agent's lexicon is determined by the role that it plays in that agent's whole set of beliefs. But, in general, no two agents' sets of beliefs are identical and no single agent's set remains invariant over time, so the meanings of two terms in the idiolects of different agents or the same agent at different times cannot be identical (pp. 8–9). Fodor and Lepore's criticism can be appreciated in graphic terms if one visualizes a simple network of wires and nodes. The holistic picture is supposed to rule that a change at any point in the system acts like an extensional displacement which affects the whole network, shifting the pos-

© *Blackwell Publishers Ltd. 1995*

ition of all the nodes in the network. If an unavoidable semantic shift afflicts all concepts of the theoretical network with every change in belief, this would render them incapable of being matched with those of the original network.

Holism does not have the consequence that every change in the beliefs held by an agent leads ineluctably to a change of meaning of all that agent's terms. In interpreting a subject, it is not the case that a disagreement with *some* of the subject's beliefs renders *all* the subject's concepts different from our own. A concept can be shared among us even though many beliefs are not shared. After several encounters with my informant, I decide that the available evidence suggests translating the term 'gavagai' by my term 'rabbit'. But I need not share all the informant's beliefs about rabbits in order to make this decision. It may turn out that the informant regards rabbits to have religious significance and that the term 'gavagai' is often mentioned in the same breath as another term which I have already translated as 'sacred'. Still, that should not force me to attribute a different concept, say the new concept *schmabbit*. I merely attribute the belief that *rabbits* are sacred. When we translate an informant's term by a term of our own, we do not expect that all the sentences in which the term appears will come out true, and we do not require agreement on all associated beliefs in order to match up a term of ours with a term of theirs. Every connection between terms need not be preserved for two terms to be correlated, as the critics of holism seem to assume.

This shows how the interpretive approach is able to escape the consequence that every change in theory leads to a change in all the concepts involved. If every single connection between terms need not be duplicated in the two theories, then every difference in beliefs will not lead to a difference in every concept. Wholesale agreement in beliefs is neither the aim of interpretation nor a necessary prerequisite. Rather than an exact isomorphism between two theories, the interpretive approach tries to achieve an overall fit. Moreover, the characteristics of this fit will be partly specified by the psychologists themselves in deciding how to interpret their protocol analyses. It may still be asked: How much agreement is necessary before we can ascribe a concept, and which beliefs are the essential ones for each concept? These are the wrong questions to ask, since the answer will generally be different for each subject being interpreted and will depend holistically on other concepts ascribed to that subject. That is not to say that the answer is radically contextual, or that subjects who are ascribed the concept *rabbit* or *skunk* on one occasion can (consistently) be withheld those concepts on another (assuming no beliefs have changed). Since an entire mental life is presupposed even when a limited number of concepts are being explored in a restricted context, only a complete characterization of a subject's states will tell us with certainty which concepts that subject has. But that is an idealization which serves as a reminder that fragmentary accounts may be misleading and may need to be revised.

© *Blackwell Publishers Ltd. 1995*

5. Conclusion

When one comes to compare a theoretical account of concepts with a prototypical account, it is apparent that they emerge from different experimental set-ups, which target different kinds of cognitive tasks. Typically, the first theory of concepts is implicated when subjects are asked to classify images or words under time pressure or to list instances of a concept in the order in which they occur to them, while the second theory is invoked when experimenters solicit detailed justifications or explanations of categorization judgments which have been made without temporal constraints. Understanding narratives, justifying classificatory judgments, and deriving complex inferential correlations are cognitive tasks that require an appeal to background theories and can be characterized in propositional terms. By contrast, tasks such as attribute-listing, word-matching, and ones in which reaction times are measured to displayed words or images, are not as conducive to an intentional characterization at all, at least not in ascriptive that-clauses. It may be possible to consider prototypes as default theories or it may be more plausible to say that these two theories regard concepts from different perspectives or stances, design and intentional.

On the intentional stance, concepts are ascribed in the course of an effort to make sense of a cognizer as a fully rational agent. Their ascription is subordinated to the larger task of understanding the agent, and they are components of beliefs that do not have a self-standing, independent existence. I have argued that such an intentional account can escape circularity by grounding the individuation of concepts in the interpretive process. What breaks the circularity among theories and concepts is the process of interpreting the psychological subject, the purpose of which is to render his or her concepts in our terms. Moreover, the standards and principles of interpretation will at least be partly determined by psychologists themselves in analyzing their experimental protocols. In interviewing subjects regarding their categorization decisions, they are constantly required to decide whether a certain concept is shared and which concept it is.

It might be protested that psychologists will be unwilling to accept such a view of concepts, since they surely take them more seriously, or at least more concretely, often treating them as physical entities with particular implementations in the brain. However, this may be reinterpreted as a request for an account of concepts from the design stance. It is doubtful that any particular facts can be adduced about representations in the brain from the perspective of the intentional stance. But nor should cognitive psychologists expect protocol analyses and mentalistic characterizations of their subjects' behaviour to issue in conclusions about the underlying mechanisms of human cognition, that is, about the design of the human categorizer.[10] When

[10] For a similar view, see Woodfield, 1993. Woodfield proposes to reinterpret the work of cognitive psychologists on conceptual development, suggesting that this work can be reconstrued without the practice of reifying concepts. When two psychologists are

© *Blackwell Publishers Ltd. 1995*

they attempt to emerge in overall interpretations of their subjects' mental lives and try to explain their behavior in linguistic terms it is obvious that psychologists are taking the perspective of the intentional stance towards them. When, by contrast, they measure such things as the rapidity and efficiency of their routine categorization judgements, they may be adopting the design stance. There is no reason to expect that the two stances will isolate the same entities. In a Forum on concepts, the Editors of *Mind and Language* once characterized what they called the 'pure attributionist view' of concepts as follows (Volume 4, p. 4): 'On that view, the business of interpretation is to cast a net of mental description over a mass of behaviour. The net has a structure: there are knots (nodes, as one might say) in it. And, to warrant the description, the behaviour needs to have some complexity too. Psychologists may investigate the whirrings and grindings that issue in that behaviour. But there is absolutely no reason to expect that they will find an inner structure matching the structure of the net; no reason to expect inner nodes corresponding point by point with the atoms of the mentalistic description.' These remarks are in keeping with what I have described as the intentional stance towards concepts. However, the Editors went on to say (*Mind and Language*, 4, p. 4): 'On the pure attributionist view, concepts fall outside the domain of psychological processes—outside the domain of detailed empirical investigation.' This further step is not warranted. According to the argument I have made in this paper, some psychologists view concepts precisely from the intentional stance and arrive at various empirical results about them without investigating the underlying 'whirrings and grindings'.

Department of Philosophy
University of Chicago

References

Armstrong, S.L., Gleitman, L.R. and Gleitman, H. 1983: What Some Concepts Might Not Be. *Cognition*, 13, 263–308.

Barclay, J.R., Bransford, J.D., Franks, J.J., McCarrell, N.S. and Netsch, K. 1974: Comprehension and Semantic Flexibility. *Journal of Verbal Learning and Verbal Behavior*, 15, 667–9.

Barsalou, L.W. 1982: Context-Independent and Context-Dependent Information in Concepts. *Memory and Cognition*, 10, 82–93.

Barsalou, L.W. 1993: Flexibility, Structure, and Linguistic Vagary in Concepts: Manifestations of a Compositional System of Perceptual Symbols. In A.F.

---

trying to decide whether a child has attained the concept *liquid*, they are not disagreeing whether a particular cognitive structure inside the two-year-old counts as a concept of *liquid*. Rather, 'The only question at issue between them is: what are the conditions that children must meet in order for them to be said to conceptualize liquids as liquids?' (p. 65)

© *Blackwell Publishers Ltd. 1995*

Collins, S.E. Gathercole, M.A. Conway and P.E. Morris (eds), *Theories of Memory*. Hillsdale, NJ.: Erlbaum.
Bilgrami, A. 1992: *Belief and Meaning*. Oxford: Basil Blackwell.
Carey, S. 1988: Conceptual Differences Between Children and Adults. *Mind and Language*, 3, 167–181.
Chomsky, N. 1980: *Rules and Representations*. New York: Columbia University Press.
Davidson, D. 1984: *Inquiries into Truth and Interpretation*. Oxford University Press.
Dennett, D.C. 1987: *The Intentional Stance*. Cambridge, MA.: MIT Press.
Fodor, J. 1994: Concepts: A Potboiler. *Cognition*, 50, 95–113.
Fodor, J. and Lepore, E. 1991: *Holism*. Oxford: Basil Blackwell.
Gleitman, L.R., Armstrong, S.L. and Gleitman, H. 1983: On Doubting the Concept 'Concept'. In E.K. Scholnick (ed.), *New Trends in Conceptual Representation: Challenges to Piaget's Theory?* Hillsdale, NJ: Erlbaum.
Gopnik, A. 1988: Conceptual and Semantic Development as Theory Change: The Case of Object Permanence. *Mind and Language*, 3, 197–216.
Grandy, R.E. 1973: Reference, Meaning, and Belief. *Journal of Philosophy*, 70, 439–52.
Haugeland, J. 1978: The Nature and Plausibility of Cognitivism. *Behavioral and Brain Sciences*, 2, 215–26.
Keil, F.C. 1986: The Acquisition of Natural Kind and Artifact Terms. In W. Demopoulos and A. Marras (eds), *Language Learning and Concept Acquisition: Foundational Issues*. Norwood, NJ.: Ablex.
Keil, F.C. 1989a: Spiders in the Web of Belief: The Tangled Relations Between Concepts and Theories. *Mind and Language*, 4, 43–50.
Keil, F.C. 1989b: *Concepts, Kinds, and Cognitive Development*. Cambridge, MA.: MIT Press.
Keller, D. and Kellas, G. 1978: Typicality as a Dimension of Encoding. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 78–85.
Lewis, D. 1974: Radical Interpretation. *Synthese*, 23, 331–44.
McCloskey, M.E. and Glucksberg, S. 1978: Natural Categories: Well Defined or Fuzzy Sets? *Memory and Cognition*, 6, 462–72.
Medin, D.L. and Wattenmaker, W.D. 1987: Category Cohesiveness, Theories, and Cognitive Archeology. In U. Neisser (ed.), *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. Cambridge University Press.
Murphy, G.L. and Medin, D.L. 1985: The Role of Theories in Conceptual Coherence. *Psychological Review*, 92, 289–316.
Neisser, U. 1987: Introduction: The Ecological and Intellectual Bases of Categorization. In U. Neisser (ed.), *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. Cambridge University Press.
Putnam, H. 1975: The Refutation of Conventionalism. In *Mind, Language, and Reality: Philosophical Papers Volume 2*. Cambridge University Press.
Quine, W.V. 1960: *Word and Object*. Cambridge, MA.: MIT Press.
Rips, L.J. 1975: Inductive Judgments about Natural Categories. *Journal of Verbal Learning and Verbal Behavior*, 14, 665–81.
Rips, L.J. and Collins, A. 1993: Categories and Resemblance. *Journal of Experimental Psychology: General*, 4, 468–86.
Rosch, E. 1978: Principles of Categorization. In E. Rosch and B. Lloyd (eds), *Cognition and Categorization*. Hillsdale, NJ.: Erlbaum.

© *Blackwell Publishers Ltd. 1995*

Rosch, E. and Mervis, C.B. 1975: Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology*, 7, 573–605.
Smith, E.E. and Medin, D.L. 1981: *Categories and Concepts*. Cambridge, MA.: Harvard University Press.
Woodfield, A. 1993: Do Your Concepts Develop? In C. Hookway and D. Peterson (eds), *Philosophy and Cognitive Science*. Cambridge University Press.

© *Blackwell Publishers Ltd. 1995*