and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory*, **7**, 418–439.
Malt, B. C., & Smith, E. E. (1982). The role of familiarity in determining typicality. *Memory and Cognition*, **10**, 69–75.
McCall, R. B., & Appelbaum, M. I. (1973). Bias in the analysis of repeated measure designs: some alternative approaches. *Child Development*, **44**, 401–415.
McCloskey, M., & Glucksberg, S. (1978). Natural categories: well-defined or fuzzy sets? *Memory and Cognition*, **6**, 462–472.
Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207–238.
Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **10**, 333–352.
Medin, D. L., & Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory*, **7**, 241–253.
Neumann, P. G. (1977). Visual prototype formation with discontinuous representation of dimensions of variability. *Memory and Cognition*, **5**, 187–197.
Nosofsky, R. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **10**, 104–114.
Nosofsky, R. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **13**, 87–108.
Nosofsky, R. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **14**, 54–65.
Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, **77**, 353–363.
Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, **83**, 304–308.
Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, **3**, 382–407.
Rosch, E. H. (1975a). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, **104**, 192–233.
Rosch, E. H. (1975b). The nature of mental codes for color categories. *Journal of Experimental Psychology: Human Perception and Performance*, **1**, 303–322.
Rosch, E., & Mervis, C. B. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology*, **7**, 573–605.
Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, **8**, 382–439.
Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, **2**, 491–502.
Tversky, A., & Kahneman, D. (1973). Availability: a heuristic for judging frequency and probability. *Cognitive Psychology*, **5**, 207–232.
Vandierendonck, A. (1984a). Concept learning, memory and transfer of dot pattern categories. *Acta Psychologica*, **55**, 71–88.
Vandierendonck, A. (1984b). Evidence for two levels of processing in concept learning and abstraction? *Cahiers de Psychologie Cognitive*, **4**, 217–244.
Vandierendonck, A. (1988). Typicality gradient in well-defined artificial categories. *Acta Psychologica*, **69**, 61–81.
Wittgenstein, L. (1953). *Philosophical investigations*. New York: Macmillan.

4

Localist and Globalist Approaches to Concepts*

KAREN LYON AND NICK CHATER†
University of Edinburgh

LOCALIST THEORIES OF CONCEPTS

A wide range of current theories of concepts all have the implicit theoretical assumption that concepts are psychologically localized. That is, a concept corresponds to a unit which *contains* the information necessary to understand that concept (Collins and Quillian, 1969; Glass and Holyoak, 1975; Katz and Postal, 1964; Smith, Shoben and Rips, 1974; Miller and Johnson-Laird, 1976; McCloskey and Glucksberg, 1979; Fodor, 1981; Barsalou, 1982). For instance, many theorists ascribe to each lexical item of natural language an entry in a mental lexicon, which must be accessed to understand the word. Thus, to have a concept is to have the appropriate 'packet' of information. Intuitively it seems that the meaning of a phrase (e.g. 'black cat') is typically derived from the meaning of its parts (the words 'black' and 'cat'). Correspondingly, complex concepts (e.g. BLACK CAT) are derived from *their* constituent concepts (BLACK and CAT). According to the localist view, concept combination is a matter of putting simple packets together. Many theorists hold such a compositional view of meaning. However, there appears to be no satisfactory account of how constituent prototypes can be combined compositionally to produce complex prototypes. We shall argue that the implicit localism of prototype views of concepts must be rejected in order to develop a theory which can adequately account for the flexibility and context sensitivity of prototypes, and of prototype combination.

* This research was funded by an SERC award (no. 85310119) to Karen Lyon and an ESRC award (no. C00428622023) to Nick Chater. We gratefully acknowledge the help of James Hampton in providing his original stimuli. Thinks are also due to Nick Braisby, Bradley Franks and Scott McGlashan, members of the Concept Discussion Group.
† The authors are joint principal authors.

Lines of Thinking, Volume 1 Edited by K. J. Gilhooly, M. T. G. Keane, R. H. Logie and G. Erdos

We shall discuss whether a definitional view which appears to be able to capture the apparent compositionality of phrases, such as 'black cat', can provide an adequate localist explanation of concept combination. We shall consider some experimental evidence that suggests that subjects' membership judgements are non-compositional, even for simple conjunctions. We also discuss theoretical objections to the definitional view, and conclude that it alone cannot provide a complete theory of concepts. We conclude that a localist view of concepts cannot be maintained and propose that a globalist alternative, in which prototype effects are knowledge-driven expectations, must be developed. In this view, concepts are a function of high-level knowledge of the world, computed on the fly.

THEORIES OF PROTOTYPE EFFECTS

Theories of prototypes were developed to explain a large body of experimental findings which suggests that some instances of a concept can be better instances than others. Subjects find rating category instances for goodness of example to be a meaningful task, and produce consistent results across a population. Moreover, these judgements correlate with performance on a variety of cognitive tasks such as resolution of anaphoric reference (Garrod and Sanford, 1977), membership verification (Rosch, 1975a, b), generation of category instances (Uyeda and Mandler, 1980) and so on. Graded membership is incompatible with a definitional view of concepts, according to which membership must be all or none. The theoretical views derived to explain these prototype phenomena are of two kinds, exemplar and probabilistic.

The Exemplar View: Instances as Localist Concept Representations

According to the exemplar view, categorization is achieved by comparing the instance in question with the stored prototype, or exemplar, of that category. The greater the similarity between the prototype and the instance, the more easily that instance can be classified as a member of the category (Rosch, 1973). Prototypicality effects are held to be analogous to effects of stimulus variation in perceptual categorization judgements; thus the exemplar view can be thought of as an extension of template-matching models in perception. Hence, effects such as those found by Posner and Keele (1968) in letter perception are assimilated into the exemplar view of categorization. Exemplars are fully specified instances, rather than abstract general category descriptions. It is not necessary that the category representation is a single exemplar—it may consist of a small group of exemplars. For example, the representation of FRUIT might consist of the exemplars of APPLE, ORANGE and PEAR.

The exemplar view localizes the meaning of a concept as a mental representation of a specific exemplar, or exemplars, of that concept. It is because of this that concept combination is wholly problematic for the exemplar theorist and no models have been put forward. An exemplar theorist would have to propose that to combine two concepts it is necessary to combine two exemplars. Thus, to understand a complex concept such as WEAPON WHICH IS ALSO A TOOL it would be necessary to combine the prototypical exemplar for WEAPON (say, GUN) with the exemplar for TOOL (say, HAMMER). As it seems impossible to combine GUN and HAMMER in order to produce an appropriate exemplar for the conjunct, it is necessary to postulate that apparently complex concepts must be represented independently (say, by an exemplar AXE). It seems that for the exemplar theorist, concepts are completely non-compositional. This, in turn, implies that it is impossible to understand phrases which correspond to novel concepts. Since we can understand and produce such novel phrases as 'blue goose' without difficulty, this comes close to a *reductio ad absurdum* of the exemplar view, as a complete account of concepts.

Probabilistic Views: Concepts as Localized Regions in Feature and Metric Spaces

According to the probabilistic views, categorization of an instance is accomplished by exceeding a critical probabilistic similarity weighting. The probabilistic views are a weaker form of the definitional view (see below), in that each property is probably, but not necessarily, true of each of its instances. The probabilistic approaches differ from the exemplar approach in that the concepts are represented by general category descriptions rather than specific instances. (cf. Kahneman and Miller, 1986, who propose a hybrid model, in which norms, which are their analogues to probabilistic prototypes, are computed on the fly from stored exemplars). The representation of a category may be more or less similar to the representations of its instances, thus generating the prototype effects. Probabilistic views can be based on either features or dimensions. We shall consider these in turn.

In the featural approach (Collins and Loftus, 1975; Goldman and Homa, 1977), a category is represented by a set of features which have a high probability of occurring in instances of the concept. The set of feature values consititutes the localist representation of each concept as summary description of the category. Hence, concept combination is a matter of combining the sets of features of the constituent concepts according to combination rules. The combination rules are typically borrowed from Zadeh's (1965) fuzzy set theory (Lakoff, 1973; Oden, 1977). This is an approach which has been heavily criticized by Osherson and Smith (1981) and by Roth and Mervis (1983).

The dimensional approach replaces discrete-valued features with continuous values along dimensions such as size. It is argued that concepts containing the same relevant dimensions can be represented in a multidimensional metric space (Shepard, 1962, 1974; Carrol and Wish, 1974; Homa and Silver, 1976). A concept corresponds to a point in such a space and so this is *literally* a localist view. The dimensions of the space are chosen to model similarity judgements about instances, such that nearby points in the space correspond to similar concepts. If an object is sufficiently near to the point corresponding to a prototype, then it is judged to be an instance of that concept. As similarity judgements have been shown to violate the basic axioms of metric spaces (Beals, Krantz and Tversky, 1968; Tversky and Krantz, 1970; Arnold, 1971; Boyd, 1972; Krantz and Tversky, 1975; Tversky and Gati, 1978; Roth and Mervis, 1983) they

are inappropriate as models of concepts. For example, the 'distance' between any two concepts need not be symmetrical. A zebra may be judged to be more similar to a horse than a horse is judged to be similar to a zebra. Eleanor Rosch (1975b) uses this phenomenon to motivate her theory that salient concepts such as HORSE act as 'cognitive reference points'.

It is unclear, in any case, how two constituent concepts, represented by points in possibly distinct multidimensional metric spaces, can be combined to generate a point in a (third) metric space representing the composite concept. No models of this process have been proposed. Finally, Smith and Medin (1981) have pointed out that dimensions can always be recoded as features. Thus, the dimensional approach can be seen as a special case of the featural approach.

Concluding Remarks on Explanations of Prototype Effects

As we have pointed out above, theories of prototype effects seem incapable of providing an account of how prototypes combine. The fact that we can understand novel sentences appears to argue for a compositional account of their meanings in terms of the meanings of their parts. Thus, to understand novel phrases such as 'blue goose' we must construct its meaning from the meaning of its components 'blue' and 'goose', rather than directly retrieving the meaning of the conjunct from memory. It may be argued that the productivity of language must be mirrored by the productivity of thought. An obvious solution to this is to construct thoughts out of stable, isolatable concepts just as in language we construct sentences out of stable, isolatable words. In this view, concepts are said to be these building blocks of thought. Prototype views fail to provide an inkling of how concepts may be combined. Hence, it seems unacceptable to identify concepts with prototype representations. To capture compositionality it appears necessary to invoke the definitional view of the structure of concepts, and to explain prototype effects as side effects of processing rather than as functions of the representation *per se*. We consider this account below.

A DEFINITIONAL VIEW: DEFINITIONS AS LOCALIST CONCEPT REPRESENTATIONS

It has been forcefully argued (Fodor, 1975, 1987; Pylyshyn 1980, 1984) that representations are solely a property of languages, whether these be natural languages or artificial languages such as logical and computer languages. Since people are able to think about the world (for example, to believe that it is raining) it is argued that there must be an internal language in which to represent such thoughts. Understanding a sentence is characterized as a process of translation from natural language into the 'language of thought'. In particular, to understand the phrase '...is a dog' involves translating it into some synonymous formula in the language of thought. Since there are an unbounded number of natural language sentences, each of which must be translated into the language of thought, the compositional structure of natural language must be paralleled

by the compositional structure of the internal language. So according to this view, the *concept* GAME WHICH IS ALSO A SPORT should be composed of the constituents GAME and SPORT (using a simple logical combination rule), just as the meaning of the *phrase* 'game which is also a sport' is related to the meaning of the words 'game' and 'sport'. Thus, the primitive concepts are combined to provide a definition of the complex concept.

The definitional view implies that the meaning of a concept is a function of the parts out of which it is composed. Primitive concepts of the language of thought are supposed to contain meaning in just the same way as it is often supposed that words of natural language do. Thus, it is argued that complex concepts inherit their meanings from the meanings of their parts, and the way they are put together, in just the same way as phrases of natural language inherit their meaning from the meanings of the words that compose them, and the way they are put together.

The definitional view typically sees the language of thought as the domain of mental operations. Sentences of the language of thought are held to be constructed from concepts. So, I can only think the thought 'This is a red table' because I have the concepts RED and TABLE, and I can put them together to create RED TABLE. Thus, concepts are used as the very building-blocks of cognition. The crucial question for the definitional theorist is: what concepts are defined in terms of other concepts (what concepts are composite), and what concepts are atomic (what are the elements of the primitive conceptual basis) (Fodor, 1981)? The traditional empiricist answer to this question is that only sensory concepts are primitive, and all other concepts are built out of these (Hume, 1975; logical positivists—e.g. Carnap, 1961; modern behaviourists—e.g. Skinner, 1957). On the other hand, rationalists such as Descartes hold that many concepts are not constructed out of sensory experience. As Fodor (1975) has argued, there can, in principle, be no theory of learning for primitive concepts, and so primitive concepts must be innate. Hence the definitional theorist must hope that all concepts can be constructed from a small set of innate primitives. To learn a new concept is merely a matter of combining these elements in a new way. It follows from this that if there is a small conceptual basis then most everyday concepts must be definable in terms of simpler concepts. Miller and Johnson-Laird (1976) and Schank and Abelson (1977) have attempted to provide such definitions. Unfortunately, there are good reasons to suppose that such a project must fail. It has been widely assumed (Fodor, 1981; cf. Chater, 1986) that if there are good definitions in the language of thought then there should also be corresponding good definitions in natural language. As, in fact, there appear to be no such good definitions in natural language it is necessary to assume that almost every word of natural language corresponds to a primitive (Fodor *et al.*, 1980; Fodor, 1975, 1981). Hence the word 'car' must correspond to a primitive, and thus innate, concept CAR. This is close to a *reductio ad absurdum* of the definitional view.

Further, there is experimental evidence (Hampton, 1988) which appears to show that concept combination is not in fact strictly compositional, even in simple cases. We shall consider this evidence in some detail.

AN EMPIRICAL INVESTIGATION OF CONCEPT COMBINATION

Concept combination can be construed as a particular sort of context effect. Each constituent of the compound acts as a context for the other through the mediation of world knowledge. For example, in understanding the complex concept PET FISH, the word 'pet' influences the kind of fish we expect, and the word 'fish' influences the kind of pet we expect, because of what we know about fish and pets. We shall argue below that context effects pose a general problem for localist theories of concepts. Even in this most simple case, of one word acting as context for another, the meaning of the whole does not seem to be a straightforward combination of the meanings of the parts.

The standard compositional account of such complex concepts (for example, in semantic theories as diverse as Montegue semantics (Dowty, Wall and Peters, 1981) and discourse representation theory (Kamp, 1981)) is a 'Boolean' model. The Boolean hypothesis is that such a semantic model is an appropriate model of how people actually make membership decisions. The hypothesis is that the members of the constituents are set-theoretically conjoined to form the members of the conjunctive concept. That is, membership of the complex concept is a Boolean function of membership of the constituent concepts. The Boolean hypothesis predicts that an item is judged to be a SPORT WHICH IS ALSO A GAME if and only if it is judged to be both a SPORT and a GAME. Hampton (1988) seeks to disconfirm the Boolean hypothesis and thus to show that concept combination is not purely compositional.

Hampton had subjects assess the membership and relatedness of lists of items to categories (stage 1) using a seven-point scale from +3 to −3. A week later subjects assessed the membership and relatedness of the same items to a conjunction of two of the categories (stage 2). For example, at stage 1, a subject might decide that chess was a good member of the category GAME (+3). However, they might decide it was not a member of the category SPORT, although it was related to it (−1). At stage 2, the subject might decide that chess is not a SPORT WHICH IS ALSO A GAME although it is related to the category (−1). This triple of membership judgements is represented (+ − −).

A difficulty with this methodology is that the rating scale confounds membership and typicality judgements. Thus the 'unitary hypothesis'—that membership and typicality judgements are 'attributed to a single underlying factor' (Hampton, 1988, p. 13)—is built into the rating scale. The unitary hypothesis contrasts with the binary view which assumes membership and typicality are distinct phenomena. It is our contention that the rating scale is inappropriate for the following reasons.

First, we need to have independent measures of both typicality and membership in order to assess whether or not they are aspects of the same phenomenon. Unless this were so we might expect the membership task to distort the typicality ratings, and the typicality task to distort the membership ratings. For example, subjects might want to rate 'tomato' as a typical VEGETABLE, although the subject knows it is a FRUIT. If membership takes precedence, then

the subject is forced to give a low typicality rating. Alternatively, if typicality takes precedence, then the subject is forced to rate 'tomato' as a VEGETABLE.

Secondly, whereas the previous point relates to judging constituents, a different problem emerges in judging membership of conjuncts. In this case, subjects might feel unhappy about giving a negative rating if the item was a good member of one of the conjunct categories. For example, chess might be judged not to be a member of SPORT. However, since it is such a good member of GAME it might be included in the category SPORT WHICH IS ALSO A GAME. Subjects may overextend their categories because they wish to express that an item which is a member of one constituent category is a better member of the conjunct than an item which is not a member of either constituent category (such as watching television). To obviate this possible spurious non-Boolean effect it is necessary to introduce a wider range of response options: A GAME WHICH IS ALSO A SPORT, JUST A GAME, JUST A SPORT, or NEITHER A GAME NOR A SPORT. Hence, subjects may judge chess to be JUST A GAME.

The following series of experiments (Chater, Lyon and Myers, 1989) is designed to clarify the status of the Boolean hypothesis.

Eight subjects participated in experiment 1, which appropriately replicates Hampton's (1988) experiment described above for comparison with the other experiments in this series. In each of the experiments the subjects were presented with lists of 16 items, for each of the six category pairs: Machine–Vehicle, Furniture–Household Appliance, Bird–Pet, Building–Dwelling, Food–Plant and Weapon–Tool. The very existence of overextensions and underextensions does not necessarily violate the Boolean hypothesis. Subjects may forget their response at stage 1 and/or change their mind at stage 2. To take this into account, Hampton uses a probabilistic model to predict the number of each type of overextension (+ − +, − + +, − − +) and the number of underextensions (+ + −) expected if subjects use a purely Boolean strategy. From Hampton's probabilistic model we derived a new prediction of the Boolean hypothesis: that, if subjects use a purely Boolean strategy, the number of overextensions and underextensions should be equal, irrespective of the rates of forgetting and changing one's mind. This is the basic measure we utilize to assess the degree of non-Boolean combination. In experiment 1, we find that subjects prefer to overextend rather than underextend their concepts ($p < 0.05$; $w(5) = 0$). This replicates Hampton's findings that subjects exhibit non-Boolean responses to membership decisions in conjuncts. But, as we have argued, the cause of non-Boolean pattern of results is unclear.

Sixteen subjects took part in experiment 2. In this experiment subjects were only required to make simple yes/no membership judgements. This should eliminate any non-Boolean effects due to the confounding of membership with typicality. A second condition provided a direct estimate of spurious non-Boolean effects: this group's responses at stage 2 were logically conjoined, after the experiment, to yield response triples comparable to those in the experimental group. Suppose that a subject rated an item as a member of both constituent categories in stage 1 (+ +), then rated the same item as a member of just one of the constituents in stage 2 (+ −). A Boolean model of the conjunction would

imply that the subject would say that the item was not a member of the conjunct in stage 2 (−). This is because, according to the Boolean hypothesis, an item is judged to be a member of the conjunct if and only if it is judged to be a member of both constituents. Thus, the response triple (constituent 1, constituent 2, conjunction) would be (+ + −). This procedure provided a direct experimental estimate of over- and underextensions due to forgetting and changing one's mind. In accordance with the prediction that we derived from Hampton's probabilistic model, there was no difference between the number of over- and underextensions in the control condition. By contrast, the experimental condition had significantly more overextensions than the control (overextensions 12%, underextensions 3.1% $\chi^2 = 11.5$ (1), $p < 0.05$). The number of overextensions was greatly reduced from experiment 1 (15.8% versus 12%). As predicted, the conflation of category membership and prototypicality exaggerated the difference between over- and underextensions (i.e. made the results less Boolean). However, there is a strong residual difference once this conflation has been removed. As mentioned above, a remaining source of error may be the restricted range of responses available to the subjects.

Thus in experiment 3 a wider range of response options is introduced at stage 2, where four options are given:

1) is a VEHICLE WHICH IS ALSO A MACHINE
2) is just a VEHICLE
3) is just a MACHINE
4) is neither a VEHICLE nor a MACHINE.

Eight subjects participated in experiment 3. We found that giving the subjects more choice did not significantly reduce the disparity between overextension (10.1%) and underextensions (3.4%) from that in experiment 2. Also, a significant difference still remained between experiment 3 and the control ($\chi^2 = 6.68$ (1), $p < 0.05$). Hence, there is a large residual non-Boolean effect. The disparity between overextensions and underextensions cannot be attributed to the lack of response choices. Comparison with the control shows that there is a real non-Boolean component to membership judgements for conjunctions. From these experiments we conclude that the tendency of subjects to overextend their category judgements is a genuine phenomenon and cannot be attributed to (i) the biasing effect of typicality judgements, (ii) the effects of forgetting and changing one's mind, or (iii) a restricted range of response options available.

Why is it that subjects are so strongly biased in favour of overextending their categories? The way in which conjunctive categories are actually used in everyday life suggests that subjects are more lenient in their membership judgements the more factors they have to take into account. For example, suppose that you have the task of casting the lead role in a local play. You are looking for a tall, handsome, blond male with a convincing Scottish accent, and a retentive memory. Given that such people are rather rare you will, of course, compromise on certain dimensions—you might settle for a sandy haired man who was not particularly tall. This amounts to overextending the category. You would be especially likely to do this if he fulfilled the rest of your criteria. This is

compensation between dimensions. In finding the best fit for a list of properties, items narrowly failing on one component are not usually excluded. This is a possible explanation of why there is a tendency to overextend rather than underextend. This is further evidence for the desirability of representational flexibility (Duncker, 1945; Ortony, 1979; Rumelhart, 1979; Barsalou, 1982). The fact that concepts do not combine according to a Boolean model is consonant with our view that world knowledge is crucially implicated in the combination of concepts.

LOCALISM AND GLOBALISM

We claim that one reason why it has been so hard to develop a theory of concepts capable of handling concept combination is that many current theories are implicitly localist. That is, the problem of concept combination is thought to be that of providing a function which somehow composes the packets of information corresponding to the constituents into a packet of information corresponding to the complex concept. We claim that the problem of concept combination is in fact a special case of the more general phenomenon of contextual influence on concepts. Just as context effects are generally agreed to depend upon world knowledge, we argue that world knowledge is crucially implicated in concept combination. We shall consider below evidence against the localist notion of concepts.

A localist view is intuitively appealing—when we say that a child has the concept DOG it sounds as if that child possesses some fixed mental structure (the concept DOG), which the child assesses when dogs are recognized. Once the structure is in place, it can be used to construct memories about dogs, generate and understand sentences involving the word 'dog' and so on. Thus the localized concept provides a building-block for the mental operations of recognition, memory and language. However, the localist view has difficulty in accounting for a wide range of well-known data. In particular it is hard to explain the facts that (i) concepts are highly flexible and context sensitive; and that (ii) as we have just shown, concepts do not appear to combine compositionally. We propose that it is the localized notion of concepts which makes these ubiquitous phenomena so difficult to explain.

In order to explain prototype effects some notion of typicality, or similarity to a prototype is used. This is problematic for the localist because the prototype and the notion of similarity used appear to be a function of the context. Thus, at the fishmonger my prototypical fish might be close to a cod; and my similarity judgements might be based on flavour, appearance and price. On the other hand, at the pet shop my prototypical fish might be near a goldfish; and my similarity metric might be based on size, shape and living conditions. There only seem to be two explanations for this phenomenon which appear to be available to the localist. The first is that these uses of the term 'fish' are simply polysemous: that is that these uses correspond to distinct lexical entries in the same way as (river) bank and (high street) bank. That lexical items have a finite set of distinct senses in the mental lexicon is the 'sense selection assumption' (Clark, 1983). However, polysemy is an inadequate explanation of context

sensitivity since there appear to be as many goodness-of-fit distributions as there are contexts (Roth and Shoben, 1983). Further it appears that prototypicality effects can be obtained for completely novel concepts like people from London with green hair (Barsalou, 1983). It seems most plausible that rather than retrieving stored goodness-of-fit distributions from memory we create them on-line. Indeed, the goodness of example distribution seems to be a reflection of what the subjects expect an instance to be like, given their knowledge of the world. Hence prototypes are a function of the whole cognitive process, rather than being building blocks out of which such processes can be constructed. This is the globalist view.

The other explanation of the flexibility and context dependence of prototypicality, open to the localist, is that changes in the goodness of example distribution are a result of post-access inferences. That is, the same conceptual representation is retrieved on all occasions of use, but subsequent inferences then determine the contextual plausibility of an instance (Gumenik, 1979; Whitney and Kellas, 1984). Whitney and Kellas (1984) claimed to provide empirical support for this point of view, by using a Stroop paradigm. Subjects were presented with a sentence which they had to later remember. Following each sentence a word was displayed in one of several colours. The subject's task was to name the colour of the word. The classical Stroop effect is that subjects find it harder to name the colour in which a word is printed if the word itself is another colour name. As subjects automatically read the word as well as recognize the colour of the stimulus, a response competition is set up between the two lexical items. This interference effect generalizes to non-colour words, particularly when they are highly activated by, for instance, contextual priming (Warren, 1972; Conrad, 1974). Hence Stroop interference effects may be used to give direct on-line information about the contextual priming of lexical items. Whitney and Kellas (1984) found, for example, that 'bird' will prime the word 'robin' more than the word 'chicken', even in a context such as 'the guest saw the bird that roasted on the grill', which strongly suggests the less typical bird. However, Lyon (1989) did not replicate these results. Potter and Faulconer (1979) report experimental findings which appear to show immediate contextual modification in a picture recognition task, given a sentential context. This implies that contextual modification is immediate and not reliant on post-access reasoning. It is important to note that if prototypes are a function of such post-access inferences then there is no reason to suppose that the postulated underlying representation has a prototype structure. If we postulate that prototype effects can be generated on-line by a process of post-access inferences, there is no reason to postulate a stored, fixed prototype representation. A stored prototype has been divested of all explanatory power. In any case, we take it to be a central task for a theory of concepts to explain how it is that inference and world knowledge can influence the goodness of example distribution in response to context. So even if there is an underlying localist mechanism of combination which is not sensitive to context (which we take to be an unresolved question), a complete explanation of the data will require a detailed consideration of world-knowledge-rich effects. Until there is firm evidence to the contrary, parsimony surely dictates that we do not postulate a 'two tier' explanation of prototype phenomena.

Murphy and Medin (1985) have pointed out the role of our naive theories of the world in conceptual coherence. Even in the early prototype views (Rosch, 1973) it was postulated that the prototype is a function of the way in which the structure of the world is described within a particular culture. This shared cultural knowledge was held to determine the prototype that an individual must have in order to possess a concept. Thus the theories by which we understand the world are crucially implicated in determining the structure of prototypes. According to this picture, our world knowledge is distilled into a localizable packet, the fixed prototype. We propose that this approach must be extended to account for subjects' rapid construction of contextually appropriate goodness of example distributions. If we hear the words 'musical instrument' at a classical concert hall our prototype might be close to violin; if we are going to a rock concert our prototype might be close to electric guitar.

The primary goal of a theory of concepts is surely to explain the everyday use of words in real contexts. Even in cases where context is underspecified (I hear the words 'musical instrument' as I turn on the radio) people attempt to provide their own contexts using their common-sense theories about what makes most sense in the circumstances (for example, what radio station it is). Thus context effects are the rule rather than the exception. We cannot view context effects as limited deformations of some fixed, stored, conceptual representation which applies in the alleged 'null' context.

We have argued that neither prototype nor definitional views appear to provide an adequate explanation of context effects. Rather than localizing the representation of the meaning of a word in some packet of information, we argue that the process of understanding is a matter of the generation of expectations on the basis of our world knowledge, rather than assessing similarity to a prototype. Let us take an analogy. Consider a landmine which 'categorizes' states of the world into two: states in which it detonates, and states in which it does not. The states of the world in which it detonates (distant earthquakes, nearby tanks, floods causing water to leak in, electric storms, and so on) need have nothing in common other than that they detonate the mine. Further, which states detonate the mine is a function of the structure of the whole mine (how shock resistant, waterproof, and resistant to electrical activity it is). The 'categorization' of the world into detonating and non-detonating circumstances cannot be explained by saying that the mine possesses some localizable structure which corresponds to the 'concept' DETONATE. Analogously, human categorization judgements cannot be understood without considering the cognitive system at large. Thus, we claim that concepts cannot be identified as psychologically localized structures but are globally determined by the cognitive system.

The problem of deploying our knowledge appropriately has been found to be crucially implicated in a wide variety of cognitive domains (Minsky, 1975). Some workers in 'artificial intelligence' have attempted to understand the way in which our theories of the world are represented using fixed primitive concepts as building blocks (Schank and Abelson, 1977; Hayes, 1978). Perhaps one reason why such approaches have been so unsuccessful (McDermott, 1986) is that they treat concepts as rigid blocks, where in fact they are as mysterious, unstable, and flexible as the processes they are used to model.

It may seem unclear how globalism can be empirically distinguished from the more traditional localist alternatives, in which concepts are held to combine compositionally, and context effects are thought to be mediated by post-access modification.

For the localist, to possess a concept is to possess a stored packet of meaning. As discussed above, the localist explains the context sensitivity and flexibility of concepts by claiming that the concept is first accessed and then modified in a contextually appropriate way. Thus, in the localist view, the application of world knowledge to produce appropriate contextual changes is seen as a high-level addition to the basic categorization process. The localist sees concepts as simple building-blocks and context effects as producing complicated modifications of these basic units. The ability to modify concepts appropriately should thus be a late development, which can only occur once the conceptual building blocks are in place. The ability to manipulate these stored concepts, producing conceptual flexibility, should gradually evolve during development. Although we should expect children's concepts to be extremely variable over time, and between individuals, we should also expect that their concepts will not be as flexible in response to contextual influence as those of adults. In order to distinguish between these alternatives it is not necessary to know what particular theories children hold, but simply whether or not children are more context sensitive than adults. Crucially, it is also unnecessary to discover whether or not different children hold the same theories, as each child will act as his or her own control. The localist predicts that children's concepts will be less context sensitive than those of adults. According to the localist, if a child possesses the concept FISH, he or she should categorize fish in the fish shop and fish in the pet shop using the same criteria (for instance: colour, surface texture, size and so on). Only later will the child learn that different notions of similarity are appropriate in different contexts, as is observed in adults (Barsalou, 1987).

In contrast, the globalist does not see the possession of a concept as a matter of having a packet of information, but rather sees the possession of a concept as a direct function of the ability to apply world knowledge. Thus categorization is directly tied to the theory of the domain in which the categorization is being made. Hence, we should expect the concept a child uses in a particular situation to depend on the theory the child has of that situation. In the fish shop, the child's concept is based on the kind of fish usually found there: cod fillets, whole trout, smoked kippers, shellfish. So, in this context, the child learns to categorize fish using criteria appropriate to the domain (such as fillets versus whole fish, smell, colour of flesh, taste). For the globalist, even at the earliest stages of development, concepts will be context sensitive, as the concepts used to make a categorization are *created* relative to the domain.

The realization that fish in the fish shop share many properties with fish in the pet shop should only arise when the child's theory is sufficiently well developed. For the child to use some of the same criteria to judge fish in the fish shop and fish in the pet shop, demands that the child's theory must be broad enough to relate together pet fish and fish for tea. Thus, for the globalist, to be able to extract uniformity across context is a high-level capacity. Globalism predicts that the abstract theories which draw out similarities between fish in different

contexts are a late cognitive development. Thus context sensitivity is the norm, and stability across contexts is a late development. The globalist prediction is that children's categorization judgements will be *more* context sensitive than those of adults. This is in direct contrast to the prediction of the localist and thus allows us to distinguish experimentally between these two opposing accounts.

A globalist theory of concepts may seem alarming since it seems that to develop a good theory of concepts we must have a theory of our common-sense understanding of the world. We have argued that concepts do not necessarily conform to localist assumptions, and that an alternative globalist framework may be preferable. Although localism and globalism are high-level theoretical distinctions, they have directly opposing predictions about the development of the context sensitivity of concepts, and can thus be distinguished empirically.

REFERENCES

Arnold, J. B. (1971). A multidimensional scaling study of semantic distance. *Journal of Experimental Psychology (Monograph)*, **90**, 349–372.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory and Cognition*, **10**, 82–93.

Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition*, **11**, 211–227.

Barsalou, L. W. (1987). The instability of graded structure: implications for the nature of concepts. In U. Neisser (ed.), *Concepts and conceptual development*. Cambridge: Cambridge University Press.

Beals, R., Krantz, D. H., & Tversky, A. (1968). Foundations of multidimensional scaling. *Psychological Review*, **75**, 127–142.

Boyd, J. P. (1972). Information distance for discrete structures. In *Multidimensional Scaling*, Vol. 1: *Theory and applications in the behavioural sciences*. New York: Seminar Press.

Carnap, R. (1967). *The logical structure of the world*. London: Routledge Kegan Paul.

Carroll, J. D., & Wish, M. (1974). Multidimensional perceptual models and measurement methods. In *Handbook of Perception*. New York: Academic Press.

Chater, N. J., (1986). The present status of the innateness controversy: a reply to Fodor (1981). Unpublished manuscript, Department of Psychology, University of Cambridge.

Chater, N. J., Lyon, K., & Myers, T. F. (1989). Why are conjunctive categories overextended? *Unpublished manuscript*. Edinburgh: University Centre for Cognitive Science.

Clark, H. H. (1983). Making sense of nonce sense. In G. B. F. d'Arcais & R. J. Jarvella (eds.), *The process of language understanding* (pp. 297–331). Chichester: John Wiley.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, **8**, 240–247.

Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, **82**, 407–428.

Conrad, C. (1974). Context effects in sentence comprehension: a study of the subjective lexicon. *Memory and Cognition*, **2**, 130–138.

Dowty, D. R., Wall, R. E., & Peters, S. (1981). *Introduction to Montague semantics*. Dordrecht: D. Reidel.

Duncker, K. (1945). On problem solving. *Psychological Monographs*, **58**, 1–113.

Fodor, J. A. (1975). *The language of thought*. New York: Thomas Crowell.

Fodor, J. A. (1981). The current status of the innateness controversy. In *Representations*. Cambridge, Mass.: MIT Press.

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, Mass.: MIT Press.

Fodor, J. A., Garrett, M. F., Walker, E. C. T., & Parkes, C. H. (1980). Against definitions. *Cognition*, **8**, 263–367.

Garrod, S., & Sanford, A. J. (1977). Interpreting anaphoric relations: the integration of semantic information while reading. *Journal of Verbal Learning and Verbal Behavior*, **16**, 77–90.
Glass, A. L., & Holyoak, K. J. (1975). Alternative conceptions of semantic memory. *Cognition*, **3**, 313–339.
Goldman, D., & Homa, D. (1977). Integrative and metric properties of abstracted information as a function of category discriminability, instance variability and experience. *Journal of Experimental Psychology: Human Learning and Memory*, **3**, 375–385.
Gumenik, W. E. (1979). The advantage of specific terms over general terms as cues for sentence recall: instantiation or retrieval? *Memory and Cognition*, **7**, 240–244.
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.
Hayes, P. (1978). The naive physics manifesto. Working Paper No. 34, ISSCO Institut Dalle Molle, Geneva, Switzerland, 1978.
Homa, D., & Silver, R. (1976). Triadic decision making in lexical memory. *Memory and Cognition*, **4**, 532–540.
Hume, D. (1975). *Enquiries concerning the human understanding and concerning the principles of morals*, 3rd edn. Oxford: Clarendon Press.
Kahneman, D., & Miller, D. T. (1986). Norm theory: conparing reality to its alternatives. *Psychological Review*, **93**, 136–153.
Kamp, H. (1981). A theory of truth and semantic representation. In J. A G. Groenendijk, T. M. Y. Janssen and M. B. J. Stokhof (eds.), *Formal methods in the study of language*, Vol. 136 (pp. 277–322). Amsterdam: Mathematical Centre Tracts.
Katz, J. J., & Postal, P. (1964). An integrated theory of linguistic descriptions. Cambridge, Mass: MIT Press.
Krantz, D. L., & Tversky, A. (1975). Similarity of rectangles: an analysis of subjective dimensions. *Journal of Mathematical Psychology*, **12**, 4–34.
Lakoff, G. P. (1973). Hedges: a study in meaning criteria and the logic of fuzzy concepts. In C. Corum, T. C. Smith-Stark and A. Weiser (eds.), *Proceedings of the ninth annual meeting of the Chicago Linguistic Society*, Chicago, 1973.
Lyon, K. (1989). PhD Thesis, University of Edinburgh.
McCloskey, M., & Glucksberg, S. (1979). Decision processes in verifying category membership statements: implications for models of semantic memory. *Cognitive Psychology*, **11**, 1–37.
McDermott, D. (1986). A critique of pure reason. Technical Report, Department of Computer Science, Yale University, June 1986.
Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge: Cambridge University Press.
Minsky, M. (1975). A framework for representing knowledge. In P. Winston (ed.), *The psychology of computer vision* (pp. 211–277). New York: McGraw Hill.
Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, **92**, 289–316.
Oden, G. C. (1977). Integration of fuzzy logical information. *Journal of Experimental Psychology: Human Perception and Performance*, **3**, 565–575.
Ortony, A. (1979). Beyond literal similarity. *Psychological Review*, **86**, 161–180.
Osherson, D. N., & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, **9**, 35–58.
Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, **77**, 353–363.
Potter, M. C., & Faulconer, B. A. (1979). Understanding noun phrases. *Journal of Verbal Learning and Verbal Behavior*, **18**, 509–521.
Pylyshyn, Z. W. (1984). *Computation and cognition: toward a foundation for cognitive science*. Montgomery, Vermont: Bradford.
Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In *Cognitive development and the acquisition of language*. New York: Academic Press.

Rosch, E. (1975a). Cognitive reference points. *Cognitive Psychology*, **7**, 532–547.
Rosch, E. (1975b). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, **104**, 192–233.
Roth, E. M., & Mervis, C. B. (1983). Fuzzy set theory and class inclusion relations in semantic categories. *Journal of Verbal Learning and Verbal Behavior*, **22**, 509–525.
Roth,E. M., & Shoben, E. J. (1983). The effect of context on the structure of categories. *Cognitive Psychology*, **15**, 346–378.
Rumelhart, D. E. (1979). Some problems with the notion of literal meanings. In A. Ortony (ed.), *Metaphor and Thought* (pp. 78–90). Cambridge: Cambridge University Press.
Schank, R. C., & Abelson, R. (1977). *Scripts, plans, goals and understanding*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
Shepard, R. N. (1962). Scaling with an unknown distance function: I. *Psychometrika*, **27**, 125–140.
Shepard, R. N. (1974). Representation of structure in similarity data: problems and prospects. *Psychometrika*, **39**, 373–421.
Skinner, B. F. (1957). *Verbal behaviour*. London: Methuen.
Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). A featural model for semantic decisions. *Psychological Review*, **81**, 214–241.
Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, Mass: Harvard University Press.
Tversky, A., & Krantz, D. L. (1970). The dimensional representation and the metric structure of similarity data. *Journal of Mathematical Psychology*, **7**, 572–597.
Tversky, A., & Gati, I. (1978). Studies of similarity. In *Cognition and categorization*. New Jersey: Hillsdale.
Uyeda, K. M., & Mandler, G. (1980). Prototypicality norms for 28 semantic categories. *Behavior Research Methods and Instrumentation*, **12**, 587–589.
Warren, R. E. (1972). Stimulus encoding and memory. *Journal of Experimental Psychology*, **94**, 90–100.
Whitney, P., & Kellas, G. (1984). Processing category terms in context: instantiation and the structure of semantic categories. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **10**, 95–103.
Zadeh, L. (1965). Fuzzy sets. *Information and control*, **8**, 338–353.