JOURNAL OF MEMORY AND LANGUAGE 34, 686–708 (1995)

Testing the Prototype Theory of Concepts

JAMES A. HAMPTON

City University, London, United Kingdom

Four experiments were designed to test two predictions of Prototype Theory. The first prediction was that when the defining (necessary) features of a concept are only partially matched by an instance, then characteristic (nonnecessary) features of concepts can affect categorization. The test of this prediction was rendered problematic as successive experiments failed to identify clearly necessary features for a range of concepts. The second hypothesis related to the independence of features in determining similarity. Most versions of Prototype Theory assume a linear combination of feature matches, which would predict that the effect of changing a feature on category membership should be greatest when the probability of categorization is closest to 50% (i.e. at the category border). The results showed that, contrary to this prediction, the effect of changing a feature was greatest when other features were all positive, and so categorization probability was at a maximum. The results support either a logistic combination rule for assessing similarity on the basis of feature match (Medin & Shaffer, 1978), or an exponential generalization function relating similarity to prototype to the sum of matching features (Shepard, 1987).

One of the most widely cited theories of concepts in psychology is Rosch's Prototype Theory (Rosch, 1973, 1975). According to Rosch, many common semantic categories like FRUIT or FURNITURE are based on concepts with a prototype structure. Possible members of such categories are categorized on the basis of how similar they are to a prototype, which is a generic representation of the common attributes of the category taken as a whole. Thus for example if the prototype for the category FRUIT were to include attributes such as *contains seeds, is sweet, grows on trees,* and *is round* (see Hampton, 1979; Rosch & Mervis, 1975, for examples), then a potential instance would be categorized as a fruit if and only if it possessed a sufficient number of these attributes (weighted for their importance). The chief characteristic of prototype concepts is that people can readily list attributes that are generally true of the category in question, but that they find it difficult to frame an explicit definition of the concept in terms of such attributes (Hampton, 1979, 1981). Instead it appears that there is a set of attributes which may carry more or less weight in the definition of the prototype, and categorization is based on whether an instance possesses enough of these attributes.

According to a recent explicit version of Prototype Theory (Hampton, 1993), the category membership of a concept such as FRUIT is determined by computing a measure of similarity to the prototype, based on degree of feature match, and by placing a threshold criterion on this feature-based similarity scale. Items above a certain higher level of similarity (that is those with enough matching features) will be clearly in the category, while those below some lower level will be clearly excluded from the category. By assuming that the placing of the

The author thanks Alison Knapp and Jean-Pierre Thibaut for practical help with the development of this research, Caroline Venables for help with data collection, and Lawrence Barsalou, Nick Braisby, Bradley Franks, John Gardiner, Helen Moss, Jean-Pierre Thibaut, anonymous reviewers, and the Chicago Concept Dining Club for helpful comments on earlier versions of this work. Address correspondence and reprint requests to James A. Hampton, Psychology Department, City University, Northampton Square, London EC1V OHB, UK.

0749-596X/95 $12.00

threshold criterion between these two limits is variable across subjects, contexts, and occasions, Prototype Theory can then readily explain the inherent fuzziness of many categories, as seen in the lack of a clear-cut category boundary. For example McCloskey and Glucksberg (1978) showed that people are inclined to disagree amongst themselves, and even to be inconsistent across occasions when asked to categorize instances of categories like DISEASE, VEHICLE, or ANIMAL, and that this "fuzziness" is largely restricted to the less typical instances of each category. Such instability in categorization can be explained in terms of variability in the placing of the category criterion and in the weight subjects may attach to different features in different contexts. A more detailed formal treatment of the theory and of how it relates to the evidence can be found in Hampton (1993).

While all accounts of Prototype Theory propose that similarity to the prototype increases with the number of matching features, many accounts remain vague on the function for combining feature matches into a similarity scale. While there has been research devoted to this topic both in the artificial concept learning literature (Medin & Shaffer, 1978; Nosofsky, 1988), and in the similarity literature (Goldstone, 1994; Goldstone, Medin, & Gentner, 1991; Markman & Gentner, 1993; Tversky, 1977; Tversky & Gati, 1982), there has been little or no attempt to model precisely how similarity relates to match of features in natural concept categorization. The most commonly assumed function (Hampton, 1979, 1993) is a weighted sum of matching features, and this is indeed the function that is most in accord with Rosch's model (see also Smith & Medin, 1981). Hampton (1979) derived a similarity measure for each instance based on summing the following function across the features for a category

$$S_j = \sum_i (w_i \cdot v_{(i,j)}) \qquad (1)$$

where w_i ($0 \leq w_i \leq 1$) is the weight of the ith feature in the prototype and corresponds to the definingness or importance of that feature for the definition of the concept, and $v_{(i,j)}$ ($-1 \leq v_{(i,j)} \leq +1$) is the degree to which the instance j possesses the feature i. Rosch and Mervis (1975) derived a similar measure of similarity to category prototype, which they called a Family Resemblance Score, based on a sum of the features that an instance possessed, weighted by how many other category members also possessed them. They showed that typicality in a category could be predicted by the Family Resemblance Score, as well as by a second measure of how many features an instance had in common with other contrasting categories. Hampton (1979) showed that the function (1) can be used in conjunction with a threshold value to discriminate members from non-members in various common categories such as BIRD or SPORT, with a reasonable degree of accuracy, even if all values of w_i are set to 1. This linear function for combining features bears a close resemblance to Tversky's Similarity model (Tversky, 1977).

One of the chief competitors to Prototype Theory is the so-called Classical Model of concepts (Smith & Medin, 1981). The Classical Model proposes that categorization of instances within a category is based on a fixed set of "defining features" which are individually necessary and jointly sufficient for categorization. Thus an instance that possesses all of the defining features must be a category member, while an instance that lacks any of the defining features cannot be a category member. Such definitions form the basis of many *taxonomic* classification systems.

The Classical Model of concepts, with roots in analytic philosophy (Sutcliffe, 1993), has been endorsed in various forms by authors who have questioned the validity of the Prototype model (Armstrong, Gleitman, & Gleitman, 1983; Landau, 1982; Margolis, 1994; Osherson & Smith, 1981,

1982; Rey, 1983; Sutcliffe, 1993). For example, Osherson and Smith (1981), considered the extent to which prototype concepts could be used as the elements in a consistent set logic, permitting definitions of set operations like conjunction, disjunction and negation. Concluding that if concepts do not have clear-cut classical definitions, then a series of logical inconsistencies and absurdities are liable to follow, they argued that the Classical Model had been rejected prematurely. Sutcliffe (1993) also presented a spirited defense of the Classical Model, arguing on a priori grounds that it is the only logically coherent account of concepts.

An apparent difficulty for the simple Classical Model is the phenomenon of instance typicality. Rosch and Mervis (1975) showed that subjects consistently rate certain category members (e.g. ROBINS) as more typical of a category such as BIRD, than other category members such as PENGUINS. Although Barsalou (1987) has shown that such typicality ratings can be remarkably unstable both across and within subjects, it is nonetheless the case that differences in the mean rated typicality of category instances produce extremely robust effects on a wide range of cognitive tasks (see Hampton, 1993 for a summary). A set of necessary features by itself clearly cannot account for such typicality differences, since all category members necessarily possess all the necessary features. This difficulty led to the proposal of a second set of features associated with any concept—termed the Characteristic Features by Smith, Shoben and Rips (1974)—which could account for differences in typicality. Thus categorization would proceed on the basis of Defining Features alone (except perhaps where rapid decisions are required), whereas differences in typicality reflect the presence or absence of Characteristic Features. PENGUINS are categorized as BIRDS because they possess the necessary Defining Features, but they are rated as *atypical* because they lack the Characteristic Features such as flight and roosting in trees.[1] Characteristic Features may also provide a means for quick identification of category members when the defining features are not readily observable. For example, the critical "defining" difference between male and female people walking down the street is not directly observable, but the characteristic features of dress, facial appearance, and body shape provide a sufficiently accurate categorization for most purposes (Miller & Johnson-Laird, 1976). The proposal that there are two kinds of feature in a concept representation—the core defining features which follow the Classical Model, augmented by a set of nondefining characteristic features determining typicality—has been termed the *Binary Model* of concept structure (Hampton, 1988, 1991, 1993), and was strongly endorsed by Osherson and Smith (1981, 1982).

In support of the Binary Model, Keil and Batterman (1984), demonstrated a developmental trend in young children, which they called the "characteristic-to-defining shift." Initially it appears that young children categorize on the basis of the most salient superficial aspects of the world—the Characteristic Features. Thus an ISLAND is a place with beaches and palm trees, while an UNCLE is a large jolly man who comes to the house bringing gifts at Christmas time. As the children grow older, Keil and Batterman (1984) documented a gradual shift in the children's definitions, to the point where they reject the use of Charac-

[1] Smith et al. (1974) introduced the distinction between Defining and Characteristic Features as being based on a *continuum* of "definingness," thus allowing that some Defining Features may not in fact be necessary for membership. However, it is clear later in their paper that they intend Defining Features to be necessary. Were they not, then the second stage of their categorization model (in which Defining Features are examined serially to ensure that they all match) would be expected to generate erroneous responses (see Hampton, 1979, for further discussion of this ambiguity in their model).

teristic Features and instead base their categorization on the less obvious Defining Features used by adults.

The Binary Model is also related to the "theory-based" account of concept structure (Murphy, 1993; Murphy & Medin, 1985), which argues for a much richer representation of concepts, taking in much of the detailed background knowledge of the world that people possess (for a developmental perspective see also Carey, 1985; Keil, 1989). Core defining features are central because of the role that they play in theoretical and inductive reasoning involving the concept. Medin and Ortony (1989), similarly suggest that the distinction of core definition and Characteristic Features serves to account for people's intuitions that the "essence" of a concept is the deep underlying reason for the observable characteristics, which thus makes a category coherent. Observable characteristic features may be "diagnostic" of category membership, but are not constitutive of the true concept definition. Medin and Ortony (1989) however appear to take a more radical view of the defining essence of a concept. They argue that the essence may in fact be an empty place-holder for many people, rather than an identifiable set of Defining Features (or other form of intensional representation) that could be made available through introspection. They propose that a *necessary* feature is not automatically an *essential* one. It may be true that all birds have feathers, so that in the Classical Model, having feathers is a defining feature of birds. However, having feathers may not be an essential feature of birds. The thesis of essentialism propounded by Medin and Ortony argues that the true definition of BIRD is in fact some deep essence, grounded in a theory of biological species, and probably involving consideration of DNA, gene pools and the like. Such essences *account* for the common features of birds, and people believe them to exist, even when they are unable to specify just what they are. This position is consistent in many respects with the "rigid designation" philosophical account of concepts (Kripke, 1972; Putnam, 1975; Rey, 1983), and serves to differentiate their "psychological essentialism" from other versions of the Binary Model in which the core definition is assumed to be a part of the concept representation possessed by the subject.

The problem with differentiating between the Binary Model and Prototype Theory is that many of the sources of evidence often used to support one or the other theory turn out on closer inspection to be inconclusive. Three sources will be considered. First, while Prototype Theory predicts differences in *typicality* amongst category members, such differences could equally well be explained in terms of differences in purely Characteristic Features (following the Binary Model), in terms of how well an instance matches some goal-derived ideal that is part of the definition (Barsalou, 1985), or even in terms of differences in frequency and familiarity amongst instances (for example in the case of some well-defined concepts, Armstrong, Gleitman & Gleitman, 1983). Second, Prototype Theory predicts the existence of *borderline cases* of category membership, but if a concept happens to have stable feature weights and/or little variability in the placement of the criterion then a prototype concept could in fact produce a clear-cut "all-or-none" category of exemplars. Alternatively, a category may appear to have no borderline cases, simply because no such cases occur in the natural world—BIRDS may be a case in point (see Hampton, 1993, for a fuller discussion). The Binary Model predicts clear-cut categorization, but it is possible that there may be other reasons for borderline cases such as lack of knowledge by the subject, or linguistic/pragmatic effects (Hampton & Dubois, 1993), which are consistent with the existence of a "core" definition. Third and finally, the Binary Model predicts the existence of features which enhance typicality *without* affecting categorization, yet the existence of such features is

not in fact incompatible with Prototype Theory. Consider the prototype concept represented in Table 1. The concept has five features, two of which have high weights, and three of which have low weights. In order to pass the criterion threshold for categorization in this case, an instance must possess a similarity to the prototype of at least 20. In effect then, the two highly weighted features are necessary for categorization and will appear to be Defining, since in the absence of either, the remaining four do not carry sufficient weight to reach threshold. Similarly, the two highly weighted features are sufficient for categorization, since any instance that possesses both features will reach threshold, *regardless* of what other features it may possess.[2] This example illustrates the point that the Binary Model may formally be reduced to a special case of the Prototype Theory—one in which each of the "Defining Features" has come to outweigh the sum of the "Characteristic Features," and where the criterion has been placed sufficiently high to require all these highly weighted features to be present.

In spite of this apparent identity between the models, there is in fact a way to differentiate between them.[3] The argument just presented only holds if the *degree* to which an instance possesses a feature is considered as all-or-none. If however degree of possession is allowed to vary (as with the parameter $v(i,j)$ in (1) above), then a discrimination between the two models becomes possible. If we suppose that an instance is being considered for the prototype concept in Table 1, but that it possesses the first feature to only a limited extent (say $+0.5$), then it is possible that possession or non-possession of the Characteristic Features f_3 to f_5 may prove important in deciding how to categorize the instance. Consider an example. Suppose that the defining features for a FACTORY are that it is a building that was designed for the purpose of manufacture, and is primarily used for that purpose. Characteristic features such as appearance, size or location would then be considered irrelevant to categorization of a building as a FACTORY. However, if a particular building did not completely fit the defining features of a factory—there was something infelicitous about its design and purpose—then it may be that the otherwise irrelevant features would now be considered as relevant. This is the prediction of Prototype Theory, because the "irrelevant" features would be needed for the item to reach the similarity criterion for membership. In contrast the Binary Model would predict that the Characteristic Features remain irrelevant to category membership in all cases, so that categorization proceeds solely on the basis of how well the instance fits the Defining Features.

The first aim of the experiments was thus to attempt to discriminate between Prototype Theory and the Binary Model, by identifying concepts with a Defining Feature which was necessary for categorization, and with a set of Characteristic Features which would not affect categorization. (Note that "Defining Feature" will be used in a theory-neutral way to refer to a feature that is *necessary* for categorization—either because it is part of some defining core, or else because it has a very high weight in the similarity function). When the Defining Feature is clearly present or clearly absent, then the catego-

TABLE 1
EXAMPLE OF A PROTOTYPE CONCEPT WITH FIVE FEATURES

Feature	Weight
f1	10
f2	10
f3	3
f4	2
f5	1

Note. Criterion for category membership is a sum of feature weights of 20.

[2] Interestingly, if the criterion is dropped to 10, then the concept takes on a *disjunctive* definition (f_1 OR f_2).

[3] I am indebted to Dan Sperber for making this suggestion.

rization probability should be unaffected by the presence or absence of the Characteristic Features (that is, after all, what makes them merely characteristic), although differences in rated typicality should be seen. However if the degree to which an instance has the Defining Feature is compromised in some way, so that categorization is unclear on the basis of the Defining Feature alone, then Prototype Theory predicts that categorization probability will be increased by the presence of the Characteristic Features. If the Binary Model is correct, then the Characteristic Features will have no effect on categorization probability, regardless of the clarity of the Defining Feature, although they should influence rated typicality when the Defining Feature is present.

The second aim of the experiments was to test an assumption of a certain class of categorization models based on Prototype Theory (Hampton, 1993). Where consideration has been given to the function relating feature overlap to similarity to prototype, most accounts propose a *linear* summation of "evidence" for categorization (see for example, (1) above). Thus the effect of changing a particular category relevant feature of an instance (for example making some hypothetical fruit instance sweet rather than bitter), should change similarity to prototype by the same amount, independently of what other features may or may not be possessed by the instance at the same time. This assumption will be termed the Feature Independence assumption.

It is important to note that Feature Independence does not translate directly into a prediction about additivity in categorization probability. Clearly if an instance possesses no other category attributes, then adding a single feature is unlikely to bring it into the category (although it may still be considered more similar to the category than it was before). The similarity will still be too far below the criterion threshold to lift the chance of a positive categorization off the floor. By the same token, adding a single feature to an instance that already has a sufficient feature match to be well above the threshold criterion will have no observable effect on the probability of categorization, since the latter will already be at ceiling, although the change will have the effect of increasing the rated typicality of the instance as a category member. According to the Prototype Theory, the relation between the similarity to prototype measure (however it is computed) and the probability of categorization follows a threshold curve as shown in Fig. 1.

As the predictions of Prototype Theory depend critically on the assumptions made about how a continuous similarity scale is transformed into a yes-or-no categorization decision, a model of this process will be briefly outlined, based on the simple notion of a threshold borrowed from psychophysics. The horizontal axis of Fig. 1 shows the range of similarity between possible instances and the concept prototype. At the left lie object classes with very little similarity to the prototype, while at the right end of the axis lies the instance with maximum similarity to the prototype. It is assumed that a categorization decision is reached by placing a variable threshold criterion on this similarity scale. If and only if similarity of an instance is greater than criterion, will a positive categorization be made. It is further assumed that the position of this criterion along the scale is variable across individuals, contexts and occasions (although not across different instances). The range of variation of the criterion is proposed to lie within the bound shown by the two vertical continuous lines on Fig. 1, while its mean position is shown by the vertical dotted line between them. Assuming that the variation in placing the criterion is normally distributed, then the probability of a positive categorization for any given level of similarity to the prototype is shown by the cumulative normal distribution function, represented by the curve in Fig. 1. When similarity is below

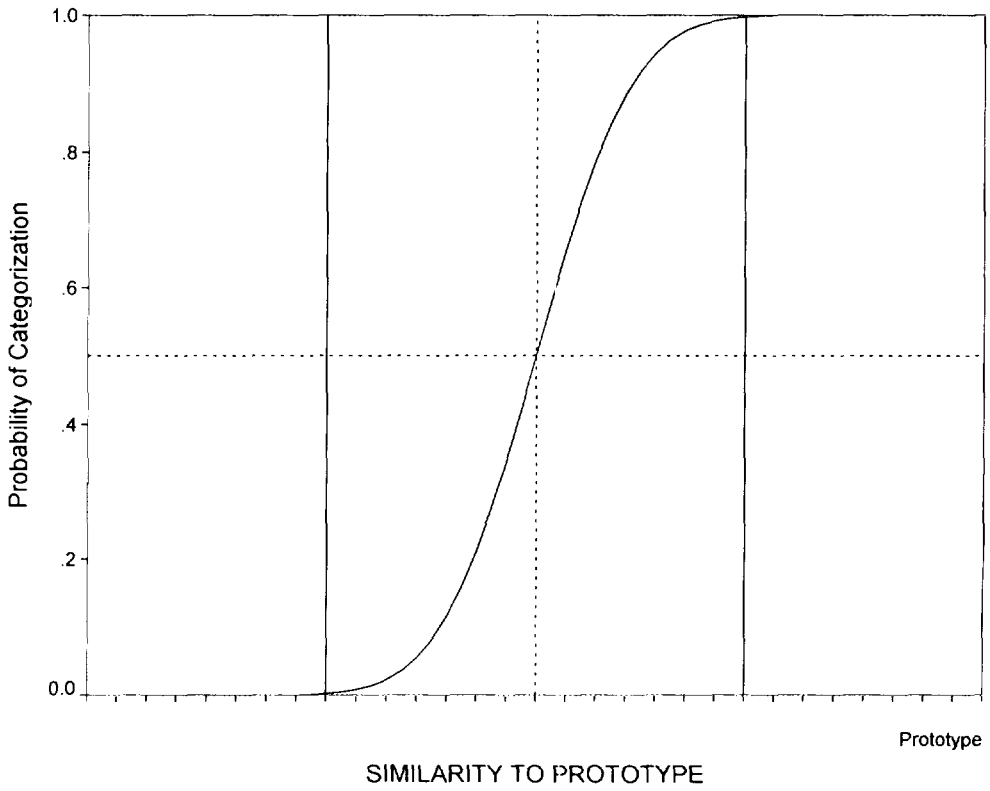

FIG. 1. Threshold curve showing the relation of positive categorization probability to similarity to prototype of some instance. The vertical solid lines indicate the borderline region in which categorization is variable across subjects and occasions.

the lower bound, then categorization is (almost) always negative. When similarity is above the upper bound, then categorization is similarly always positive, and when similarity lies between the two bounds, then the probability of categorization lies between 0 and 1—this is the "border" region where subjects disagree and are inconsistent in their classification of instances.

Independence of features in judging similarity to prototype predicts that the effect of adding (or removing) a feature from an instance will be to produce a *constant* shift to the right (or left) along the similarity scale. The relation between a change in similarity and a change in categorization probability is shown by the slope of the function in Fig. 1, and this slope is at a maximum when the categorization probability is at 50%. Feature Independence therefore predicts that the effect of altering a feature of an instance on categorization probability will be greatest when the instance has a categorization probability of 50% (more precisely, the effect will be maximal when the increase in similarity due to adding the feature takes an instance from a similarity point some distance below the 50% threshold shown by the vertical dotted line, to a point an equal distance above that threshold). In effect, this prediction falls out of the model because the sensitivity of categorization probability to changes in an instance's features must be greatest when the instance is closest to the category border (note that the category border has been *defined* here as that point where the probability of categorization is 50%). Although

this prediction follows in a straightforward manner from Prototype Theory, it has not previously been tested. The second purpose of this paper is to make a direct test of this prediction.

As a prerequisite for the study it was necessary to identify concepts that had a clearly necessary Defining Feature, and a set of Characteristic Features that would not normally affect categorization. A series of four experiments were conducted with this aim. From experiment to experiment, the concepts and features were developed and retested in a recursive fashion. As will be seen, in practice it proved extremely difficult to identify a sufficiently large set of concepts that had the desired properties. In itself this may reflect a lack of imagination on the part of the investigator. Alternatively it may reflect the very narrow range of concepts that actually fit the pattern predicted by the Binary Model. Inter-subject variation in categorization proved to be much more widespread than was expected on the basis of earlier reports (e.g. Keil and Batterman, 1984; Rips, 1989). Although the first aim of the experiments therefore proved difficult to fulfill, the resulting data provided an opportune source of evidence to test the second hypothesis—the Feature Independence assumption of Prototype Theory. For brevity's sake, the four experiments will be described together. The pooled data from the experiments will then be used to test the prediction relating to the Independence of features in determining similarity, and hence categorization probability.

EXPERIMENTS

Method

Subjects. The four experiments each employed 72 subjects, with the exception of Experiment 2, which employed 36 subjects who each judged two versions of each concept. No subject acted in more than one experiment. Subjects were adults largely drawn from students and employees of City University, London. All had British English as their native language.

Materials. An initial list of concepts was drawn up for the first experiment, based partly on materials used by Keil and Batterman (1984) who charted the developmental change in children's categorization from a classification based on surface characteristics to one based on adult definitions. A list of the concepts (26 in all), together with the experiments in which they were used, is shown in Table 2.

The development of materials from experiment to experiment proceeded as follows. For Experiment 1, a set of 18 concepts were selected, after some piloting with a larger selection of materials. For

TABLE 2
CONCEPTS USED IN THE FOUR EXPERIMENTS

Concept	Expt. 1	Expt. 2	Expt. 3	Expt. 4
Bank	(+)	+	+	–
Bird	–	–	+	–
Book	+	+	–	–
Chair	+	+	+	+
Church	–	–	–	+
Cousin	–	–	–	+
Dustbin	+	+	+	+
Factory	+	+	+	+
Gun	+	+	–	–
Holiday	+	+	+	+
Hotel	+	+	+	–
Human	–	–	–	+
Lie	+	+	+	+
Lunch	+	+	+	+
Museum	+	+	–	–
Orange	–	–	+	–
River	+	+	+	–
Road	+	+	+	–
Streetlight	+	+	+	+
Supermarket	+	+	–	–
Swimming pool	–	–	+	+
Taxi	+	+	–	–
Theft	+	+	+	–
Umbrella	+	+	+	–
Uncle	–	–	+	+
Zebra	–	–	+	–

Note. Bank was included in Experiment 1, but was omitted from the analysis because of a printing error in one of the versions. Experiments 1–3 employed 18 concepts each; Experiment 4 used 12.

each concept a defining feature (DF) was chosen, which had a prima facie case for being necessary. All other defining features were held constant and were assumed to be present in their normal form. For example, for the concept FACTORY, all scenarios described a building with a particular industrial-related function. The DF+ scenarios specified

> The interior of the building was specifically designed to contain machines which are used for making things which are then distributed and sold in shops. This is the sole function of the building.

The DF− scenarios by contrast specified

> The interior of the building was specifically designed to contain specially modified machines which are used solely for training young people in the skills of machine maintenance. This is the sole function of the building.

For STREETLIGHT the DF+ was that a light had been put up with the express intention of helping users of a street to see where they were going in the dark. The corresponding DF− was that the light was put up with the intention of lighting a factory yard as a deterrent to burglars. Note that other DF such as that the Factory was used by people at work, or that the Streetlight cast illumination were present in all scenarios.

Characteristic features (CF) were then also created, with the intention that they would be neither necessary nor sufficient (as a set). For example for FACTORY the CF+ were that the building was

> A long low building made of prefabricated concrete, with air vents along the ridge of the roof. The doors are huge allowing large trucks to enter the building.

The CF− were that the building was

> A smart red brick house, with a pretty garden in front, and ivy growing up the walls. The front door is oak panelled with a smart brass knocker.

For STREETLIGHT, the CF+ was a description of a normal looking streetlight, while the CF− was a description of a hologram projector that cunningly cast light onto the appropriate spot (the street or the yard).

In choosing DF, the commonly accepted notion was applied that biological kinds are defined by "essence" as determined by genealogical history, while artifact kinds are primarily defined by the intended function for which they were created, and subsequently used (Carey, 1985; Keil, 1986; but see also Malt, 1991). CF were largely chosen to be superficial appearance features, reflecting the most typical instances in terms of frequency and ideals (Barsalou, 1985).

Given the selection of DF and CF, it then remained to create a scenario instance that would have the DF compromised in some way, so that the instance neither clearly possessed, nor clearly did not possess the DF. In the case of FACTORY this was that

> the interior of the building was specifically designed to contain machines which are used for repairing broken objects. This is the sole function of the building.

Thus the building was involved in the production of artifacts, although not in their original manufacture. For STREETLIGHT, the compromised DF involved a story whereby the light was installed for illuminating the factory yard as a deterrent to burglars, but local residents persuaded the factory owner to install a mirror, so that the light also lit up the road. These compromised DF scenarios will be labeled as [DF(?)].

The combination of three levels of the DF (+, −, and ?) and two levels of the CF (+ and −) yielded six scenario instances for each concept. From experiment to experiment, additional piloting was run, and the materials were modified in an attempt to increase the effectiveness of the DF in relation to the CF. Ideally, the experiment requires that both DF+ scenarios receive 100% positive categorization, while both DF− scenarios receive 0% positive categorization (that is the DF is treated as a nec-

essary feature by all subjects). Each of the experiments used a set of 18 concepts, except for Experiment 4 which used just 12 in an attempt to improve the match of the data to this ideal. (Looking ahead to the results in Table 3, it can be seen that there was a steady improvement in the match of the materials to the requirement of a pattern of 100/ 100/ 0/ 0 percent positive categorization for DF+CF+, DF+CF−, DF−CF+, and DF−CF−, respectively, but that even by the fourth experiment there was still some systematic deviation from the ideal pattern.) Examination of the materials in the first experiment showed that one possible account for positive categorization of the DF− might have been that these scenarios fit no other known category. Thus a subject may have continued to say that the object belonged in the category, because it was clearly not in any other known class. Later experiments remedied this possible problem by arranging that DF− should typically belong in a contrasting category.[4] Thus, for example, the ZEBRA became a horse, or the HOTEL became an apartment block. The effect of this change was to help reduce positive responding to DF−, but not to the point where CF+DF− showed zero positive responses. Further examples of concept scenarios for the concepts UNCLE and UMBRELLA are shown in Appendix A.

Design. The design of the four experiments was identical with one exception. The design of Experiments 1, 3, and 4 employed six groups of subjects, with materials rotated across the six groups, so that each subject saw just one of the scenarios for each concept, and so that over all concepts, each subject saw an equal number of scenarios of each type. Experiment 2 employed a repeated measures design in which subjects saw each concept twice. Across the two scenarios seen by a subject for any concept, the DF was kept constant, while the CF was varied. Scenarios were again rotated across six subject groups defined according to which level of DF was used, and the order in which the two scenarios were judged.

Scenarios were printed two or more to a page, and booklets for each subject group were assembled with the order of pages approximately balanced across subjects.

Instructions. Instructions to subjects were changed slightly from experiment to experiment in an attempt to encourage subjects to treat the DF as defining. In the first experiment, subjects were simply asked to tick a box for Yes or No in response to a question such as "Is this a factory?" or "Does Harry have a holiday?" In addition they were asked to give a 5-point typicality rating for Yes responses (from Highly typical to Highly atypical), and for No responses they were asked to rate "How close does it come to being an example" on a 5-point scale from Very close to Very distant. (Typicality ratings were required so that the effectiveness of Characteristic Features could be demonstrated.) In subsequent experiments, the instructions were made stricter by emphasising to subjects that they should respond on the basis of how the category is really defined. For example, for Experiments 3 and 4, the instructions gave an instance of the concept CARPET as an example, and then included the following:

> Read the description of the example carefully, and decide whether or not this example fits your definition of what a carpet (for example) really is. . . . Base your decision as much as possible on what you think is the real definition of the term, rather than on whether one could use the word to refer to the object. For example a stone statue of a lion, is not a real lion and so you should say No if asked if it is a lion, even though one could reasonably use the word 'lion' to refer to it.

The instructions went on:

> In considering each case, you should assume that everything else about the example that has not been specifically described is normal–for instance in the example above, you could assume that the cloth was a normal thickness, that the floor was flat and that the room was in a human dwelling of some kind. If you feel that it is im-

[4] I am indebted to Jean-Pierre Thibaut for this suggested improvement.

possible to decide about a classification because the decision critically depends on some crucial information which is missing and cannot be reasonably inferred, then please write down next to the example what that information would be. (It is hoped that you will not need to use this option, as all relevant information will have been supplied.)

After the first two experiments, the typicality/closeness judgement was also removed from the initial phase of the experiment, in case it should perhaps lead subjects to adopt a "relativist" approach to categorization. In Experiment 3 and 4, subjects simply went through the booklet answering Yes or No to questions such as "Is this really a chair?" When they reached the end, a final page of instructions asked them to go back, and give a typicality rating to all the Yes responses that they had made. Closeness of No responses was not rated. Booklets were distributed to subjects who completed them in their own time.

Results

The percentage of subjects who gave positive responses to the six types of scenario is shown in Table 3, averaged across the scenarios in each experiment. (One scenario from Experiment 1 had to be dropped, owing to a misprint in one of the versions). In the following sections, each of the two predictions of Prototype Theory will be considered.

Compensation by characteristic features. The first aim of the experiments was to test a prediction of Prototype Theory that when a DF was only partially matched, then the presence or absence of a set of CF would influence categorization. It can be seen in Table 3 that the ability to test this prediction was compromised by the difficulty of finding DF that were treated as defining by all subjects. In all experiments, there was a tendency for the DF+CF− scenarios to be categorized positively less often than the DF+CF+ and for the DF−CF+ scenarios to be more often positively categorized than the DF−CF− scenarios. In Experiment 4, the effect of the CF was finally reduced to around 10%, but there was still an overall effect of the CF on categorization. In light of the categorization frequency results, mean typicality ratings will not be reported.

It appears to have been extremely difficult to identify concepts for which subjects agreed on a clear binary distinction between DF and CF. These results were indeed quite surprising on the basis of many earlier analyses of the defining "essences" of artifacts and natural kinds. Artifacts have often been seen as being primarily defined by the function for which they are used, and for which they were intended by their maker (Keil, 1986; Rips, 1989) (although see Malt, 1990, 1991, for contrary evidence). On the other hand biological kinds have been thought of as defined in terms of some presumed essence, such as a set of DNA codes, which may not even be

TABLE 3
PERCENTAGE OF SUBJECTS GIVING POSITIVE CATEGORIZATIONS TO EACH TYPE OF SCENARIO, AVERAGED ACROSS CONCEPTS FOR EACH EXPERIMENT

	DF+		DF(?)		DF−	
Experiment	DF+CF+	DF+CF−	DF(?)CF+	DF(?)CF−	DF−CF+	DF−CF−
1	96	63	68	42	52	29
2	99	66	71	44	48	22
3	92	72	52	31	21	12
4	92	78	45	37	13	8
Mean	95	70	59	38	33	18
CF effect	25		19		15	

Note. DF, Defining Feature; CF, Characteristic Feature; ?, partial match to DF.

known to most people, but which can reasonably be inferred from genetic inheritance (Carey, 1985; Rips, 1989.) Materials were constructed on this basis. Surprisingly, in the present data genetic essence did not provide a necessary and sufficient condition for biological kinds. For example in Experiment 3, the DF−CF+ scenario for the concept ORANGE was as follows:

> A round fruit which was grown from a real Lemon tree, using new high-tech special growing conditions. It has a waxy orange peel, and contains segments which contain a sharp sweet-tasting orange-flavored juice. It is pleasant to eat on its own.

Unlike Rips' (1989) results in which a bird that was metamorphosed by radioactivity into an insect was still judged to be a bird, many of the present subjects (4 out of 12) stated that the fruit was really an Orange. They were apparently quite happy for Lemon trees to bear Oranges as fruits, as a result of purely environmental conditions.

The corresponding DF+CF− scenario used the same story, except that the tree was a real Orange tree, while as a result of growing conditions the fruit was yellow and had sour-tasting lemon-flavored juice, which was unpleasant to eat on its own. In this case, only 4 out of 12 subjects judged the fruit to be really an Orange, in spite of its clear presumed essence.

Similar problems arose with another biological kind, ZEBRA. Consider the following DF+CF− description:

> The offspring of two zebras. The creature was given a special experimental nutritional diet during development. It now looks and behaves just like a horse, with a uniform brown color.

When asked if this was really a zebra, only 4 out of 12 subjects agreed, the rest of the subjects ignoring the genotype in favor of the phenotype, contrary to the assumptions of psychological essentialism (Medin & Ortony, 1989; Rips, 1989).

The overall pattern may be obscuring individual differences amongst concepts. For example UNCLE, which is explicitly definable in terms of kinship relations of siblinghood, parent-hood and marriage, showed no effect of the CF when the DF was partially matched (DF(?)) in either Experiment 3 or Experiment 4. On the other hand concepts like SWIMMING POOL (Experiments 3 and 4) and BIRD (Experiment 3) showed large CF effects for DF(?). Unfortunately, given the variability across materials and across experiments, it was not possible to test meaningfully for individual concept differences. Previous research has indicated that the way in which concepts are defined probably varies as a function of the ontological type of the concept. For example, natural biological kinds are considered to be defined by hidden essence, whereas artifacts are defined by function and use (Carey, 1985; Keil, 1986). Other nominal kinds like UNCLE or LUNCH (Keil & Batterman, 1985) may have explicit conventional definitions. The difference between biological and artifact kinds is shown up for example in the different susceptibility of objects to change their "type" when either new discoveries are made about the material from which they are made, or transformations are applied to their external appearance (Keil, 1986). These differences in concept definition were built into the selection of DF in the present experiments. Looking post hoc at natural biological kinds versus artifact kinds, there was no evidence for any differential tendency for either a better fit to the assumption that the DF were truly defining, or a stronger effect of the CF features when the DF were partially matched. The failure to find consistent results was not therefore a function of having a variety of concept types in the selection of materials.

Feature independence. The second prediction of Prototype Theory was that the effect of changing a category feature on categorization probability would be greatest across the category border (as defined by 50% categorization probability). Considered overall, the data in Table 3 suggest that the CF effect was actually *greater* when the

DF was matched (+) than when it was either partially matched (?) or not matched at all (−). An analysis of variance was carried out, using scenarios-within-experiments as cases ($N = 65$, after excluding the scenario dropped from Experiment 1), and the frequency of subjects responding Yes as the dependent variable. Type of Scenario was treated as a two-way repeated measures design with DF having three levels, and CF having two levels. Experiment was included as a between-groups factor. The analysis showed significant main effects of DF and CF, and significant two-way interactions between Experiment and DF ($F(6,122) = 6.94$, $p < .001$), and between Experiment and CF ($F(3,61) = 8.68$, $p < .001$), reflecting the gradual increase in the DF effect and gradual reduction in the CF effect as the materials were refined. There was also a significant two-way interaction between DF and CF ($F(6,122) = 4.66$, $p < .02$). The CF effect was greater when the DF was matched (25%), than when it was partially matched (19%) or not matched at all (15%). There was also a significant linear trend in the CF effect across the three levels of the DF factor.

In order to remove possible artifacts arising from the use of frequency data, the results were transformed and reanalyzed as follows. One particular factor that may be distorting the pattern of results is the fact that the analysis is based on probability of categorization, whereas it is the underlying similarity dimension that is in fact more relevant to the question of feature independence. As described in the introduction, the effect of increasing feature match (similarity) on the probability of categorization should not be linear but should rather follow a curve such as that in Fig. 1. Accordingly, the frequency data were transformed into a hypothetical similarity scale as follows.

First, frequencies of 12 out of 12 were recoded as 11.75, on the basis that they reflect a frequency somewhere between 11.5 and 12 on the scale. Similarly frequencies of 0 were recoded as 0.25. The frequencies were then divided by 12 to convert them to probabilities. These probabilities were subsequently transformed into z scores using the probability density function for the normal distribution. That is to say each probability was converted to the value of z which would leave that area in the positive tail of the normal distribution. Thus a frequency of 11.75 was transformed to a probability of 0.979 and a z of 2.04. A frequency of 6 became a probability of .5 and a z of 0.0, and a frequency of 0.25 was converted to a probability of 0.021 and a z of −2.04. The means in Table 3 were then recalculated on the basis of the z-transformed data and are shown in Table 4. Feature Independence would then predict that having removed the effect of the non-linearity result-

TABLE 4

MEAN z SCORES FOR PROBABILITY OF SUBJECTS GIVING POSITIVE CATEGORIZATIONS TO EACH TYPE OF SCENARIO (FREQUENCIES CONVERTED TO z SCORES BEFORE AVERAGING ACROSS CONCEPTS FOR EACH EXPERIMENT

	DF+		DF(?)		DF−	
Experiment	DF+CF+	DF+CF−	DF(?)CF+	DF(?)CF−	DF−CF+	DF−CF−
1	1.80	0.39	0.47	−0.22	0.02	−0.69
2	1.93	0.45	0.70	−0.20	−0.05	−0.88
3	1.53	0.74	0.11	−0.66	−0.94	−1.32
4	1.49	0.88	−0.09	−0.44	−1.25	−1.54
Mean	1.70	0.59	0.33	−0.38	−0.50	−1.08
CF effect	1.1087		0.7042		0.5732	

Note. DF, Defining Feature; CF, Characteristic Feature; ?, partial match to DF.

ing from threshold effects, the resulting pattern of data should show equal size CF effects at all three levels of the DF. Effectively the z transformation of the probabilities straightens out the threshold curve in Fig. 1.

The ANOVA was recalculated using the z transformed scores for each scenario. The pattern of results for the transformed scores was the same in all major respects as for the raw frequencies, with a significant two-way interaction between DF and CF, ($F(2,122) = 10.99, p < .001$), which did not interact significantly with Experiment ($F(6,122) = 1.07$). The pattern of means was also repeated, with a larger CF effect for the matching DF, than for either the partial matching DF or the non-matching DF. The linear trend in the CF effect across levels of DF was again significant.

Plotting the CF effect. Analysis on the basis of the DF and CF fixed effect factors may present a distorted picture because of different levels of positive responding for DF+CF+ and because of different sized CF effects occurring across different concepts. The three levels of the DF effect could correspond to different degrees of similarity (and hence of categorization probability) across different concepts. Similarly, the CF effect itself was stronger for some concepts than for others. The z-transformed data were therefore further manipulated in order to produce an unbiased scatter plot of the function relating the size of the CF effect to level on the underlying similarity scale. Appendix B gives a worked-out example to illustrate the set of data transformations undertaken. The transformations were applied to concepts within experiments, so that a total of 65 cases were available initially. Selection of a subset of 36 concepts was subsequently necessary in order to arrive at a sensitive measure of the effect of manipulating the set of Characteristic Features on categorization for different levels of similarity to the prototype (as reflected by the manipulation of the Defining Features). The transformations were as follows:

(a) The sizes of the three CF effects corresponding to each level of the DF (+, ?, and −) for each concept were determined by subtracting the z value for CF− from the z value for CF+, as in (2) to (4).

$$\text{CF effect for DF+} = z[\text{DF+CF+}] - z[\text{DF+CF−}] \quad (2)$$

$$\text{CF effect for DF?} = z[\text{DF?CF+}] - z[\text{DF?CF−}] \quad (3)$$

$$\text{CF effect for DF−} = z[\text{DF−CF−}] - z[\text{DF−CF−}] \quad (4)$$

(b) In order to equate across materials for the different effectiveness of the CF manipulations, the raw CF effects from (a) were then standardized by dividing each of the three CF effects for each concept by their mean. The three standardized CF effects for each concept therefore always had a mean of 1. The relative difference in their effectiveness as a function of the level of similarity to the concept could then be seen irrespective of the overall level of the CF effect. This procedure only makes sense if the Characteristic Features are having a positive effect on categorization. Division by the mean for those concepts where the CF were having little or no effect would simply magnify noise in the data. It was therefore necessary to restrict the analysis to those concepts where the CF were having some measurable effect. An arbitrary cut-off of 0.5 for the mean CF effect from (a) above was used to select 49 of the 65 concepts across experiments for the analysis. A further restriction on cases was necessary because of ceiling and floor effects at either end of the scale. Where both DF+CF+ and DF+CF− scenarios were attracting nearly 100% positive categorization, the scope for a CF effect is obviously artificially restricted by a ceiling effect. A similar problem of a floor effect arises where both DF−CF+ and DF−CF− are attracting nearly 0% positive categori-

zation. Both problems are caused by the difficulty of estimating similarity reliably when the categorization probability is asymptoting at either zero or one (see Fig. 1). A further selection of concepts was therefore made which excluded 10 concepts suffering from ceiling effects and 3 that suffered from floor effects. This exclusion was done by restricting the range of the average similarity to -1.5 to $+1.5$ on the scale (see the ordinate of Fig. 2). The final selection was therefore of 36 concepts out of the original set of 65. Note that the selection of these concepts for this analysis is motivated simply by the need to find concepts that on the one hand have effective Characteristic Features, and that on the other hand have Defining Features which are not either necessary or sufficient on their own. It is only for these concepts that the question of Feature Independence can be easily addressed.

(c) The standardized CF effect data were paired with the mean z values of the CF+ and CF− versions of each scenario. For example the standardized CF effect based originally on the difference between the DF+CF+ and DF+CF− scenarios was paired with the mean z value for the DF+CF+ and DF+CF− scenarios. A scatterplot was then produced of the 108 ($= 36 \times 3$) pairs of data points. Figure 2 shows the size of the CF effect (as a standardized difference in z scores) as the abscissa, against the mean z value of the corresponding CF+ and CF− scenarios as the ordinate.

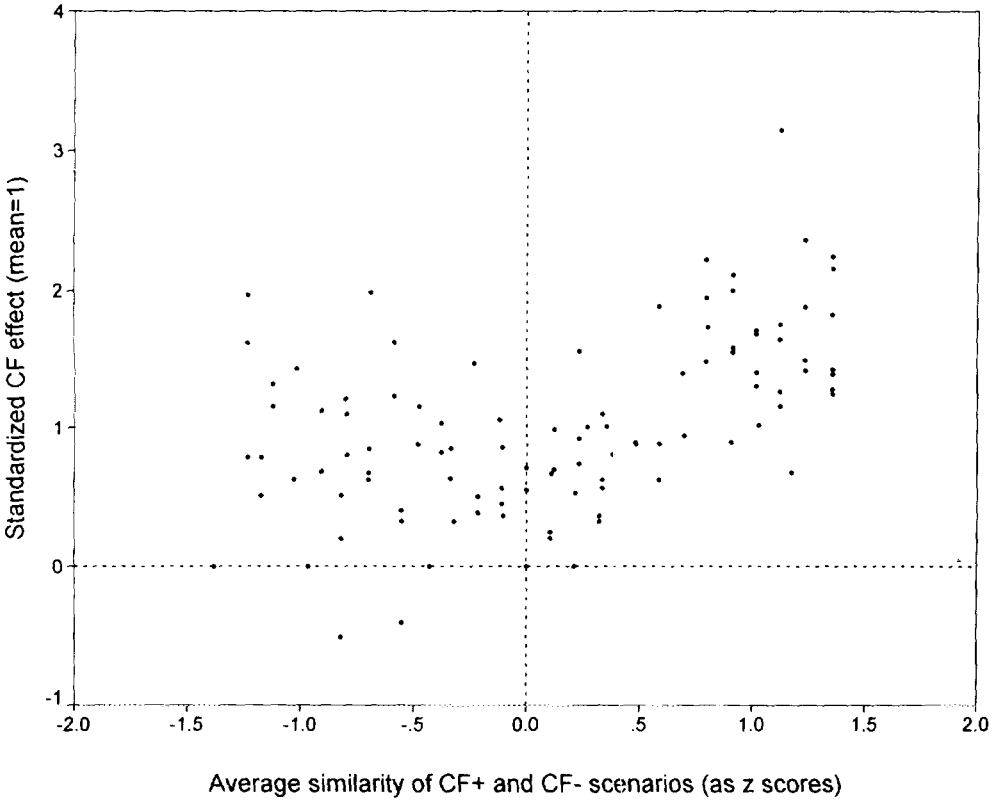

FIG. 2. Scatter plot of standardized CF effect (defined as the difference in z values between CF+ and CF− averaged to a mean of 1.0 across the three DF scenarios) plotted against the mean z value of the CF+ and CF− scenarios. Concepts with low CF effect size have been excluded. Data points beyond 1.5 at either end of the scale have also been excluded.

The figure shows a positively sloped scatter of points, confirming that the CF effect is larger the closer an instance is to the concept prototype. Overall the correlation $r(106)$ was 0.5301, $p < .001$. Note that attenuation effects due to the end of the scale have been largely eliminated by the restriction of the ordinate to a range of -1.5 to $+1.5$ (the full scale ran from -2.04 to $+2.04$). Further analysis of the data shows that the positive correlation between the CF effect and average similarity was only present in the *top* half of the scale. In a second order polynomial regression, the quadratic component of the ordinate scale was significant and positive. When correlation coefficients were calculated for the top and bottom half of the scale separately, for z values ≥ 0, $r(56)$ was .689, ($p < .001$) while for $z < 0$, $r(48)$ was $-.069$ (N.S.). This pattern shows not only that the CF effect was greater for higher levels of similarity, but also that the slope of the function relating CF effect to similarity was positively accelerated. There was no apparent change in the CF effect with decreasing similarity once the average probability of categorization for the two scenarios dropped below 50% (corresponding to $z = 0$ on the scale).

Discussion

Prototype Theory versus the Binary Model

This paper set out to provide two tests of Prototype Theory. The first test aimed to differentiate between Prototype Theory and the Binary Model. This test proved problematic because in spite of four attempts to refine the materials, instructions and procedure, the CF were still having an effect on categorization both when the DF was present (DF+) and when it was absent (DF−). It must be concluded then, that the primary assumption of the Binary Model—that there exist CF which are not involved in categorization—received no support. Contrary to the expectations of this model, the analysis of attributes into Defining versus Characteristic Features proved difficult to establish and sustain across a range of concepts. In spite of strict instructions to subjects to consider the real nature of the objects being described, and four cycles of rewriting and editing the scenarios, categorization probability continued to be sensitive to manipulation of features which prima facie would not be expected to influence category membership. Particularly noteworthy was the tendency of subjects to classify biological kinds on the basis of superficial appearance as well as on the basis of presumed genetic essence. It was therefore impossible to conduct the originally planned test of the Binary Model, because of the difficulty of finding sets of materials that conformed closely enough to the requirements of that hypothesis. It should be noted that although Prototype Theory's account of why certain features may appear necessary could not be tested, the failure to find appropriate necessary features in the experiments can also be taken as evidence for the basic assumptions of the Prototype Theory.

What interpretation can be offered for these results? At first blush it would appear that the data are at odds with earlier evidence from Keil and Batterman (1984) and Rips (1989). Keil and Batterman (1984), for example, found that older children switched from classifying according to superficial Characteristic Features in favor of classifying by Defining Features. A closer examination of Keil and Batterman's results shows that the oldest group (mean age 9 years 9 months) were successfully rejecting scenarios of the DF−CF+ kind, but were still considerably below perfect performance on correctly accepting scenarios of the DF+CF− kind. The evidence is therefore that the older children had learned about the importance of the DF, but not that they were convinced about the *unimportance* of the CF. The data presented here suggest that for many concepts, even adults are not convinced that CF should not affect categorization (Keil and Batterman did not report an adult control

group in their published account). It is also interesting that Keil and Batterman comment that when devising materials they found it necessary to restrict themselves largely to kinship terms, terms for social conventions, and financially and morally related concepts, in order to find suitable sets of DF and CF (although they also included artifact concepts like HAT and CHURCH). Keil and Batterman's data are certainly consistent with a developing prototype representation in which as knowledge increases, new features are added to the concept representation, and the threshold criterion is set sufficiently high for many children to believe that only items with both DF and CF present are category members. (It is an advantage of Prototype Theory that prototype representations very readily and simply allow for conceptual growth and change through the gradual accretion of features, and the gradual learning of their appropriate weights.)

In another often cited critique of similarity-based categorization models, Rips (1989) reported a study in which subjects were given scenarios very like those used in the experiments reported here. In one scenario a creature with all the features of a bird is transformed through living near a toxic waste dump into something that looks and acts like an insect. (The study involved a number of different animals undergoing similar accidental life-time changes). Rips reported that in such a case subjects tended to rate the transformed creature as still a bird, although it was now considered more similar to an insect. A second study examined artifacts. For example, in one scenario two umbrella-like objects were rated for category membership as umbrellas. One of the two looked like an umbrella but was designed for and used as a lampshade. The other looked like a lampshade, but was designed for and used as an umbrella. The results again differentiated similarity ratings from categorization. The majority tended to categorize according to designed function rather than appearance, whereas similarity ratings also took account of appearance. (The generality of some of Rips' findings has since been challenged by Smith and Sloman, 1994.)

On closer examination, Rips' data turn out to be less different from those reported here than might be initially supposed. When the unfortunate bird-like creature is metamorphosed into an insect, the categorization rating actually dropped from 9.5 to 6.4 on a scale of 1 = insect to 10 = bird. (Means have been estimated from the figures in Rips, 1989.) Rips did not ask subjects to make a binary categorization choice, instead opting for a "likelihood of categorization" scale, but his data are certainly consistent with the possibility that if forced to choose if the creature were still a bird, a substantial minority of subjects would have said no. Rips indeed admitted that "we were unable to change similarity ratings without changing categorization at all," a finding that is replicated in the present study. Similar results were obtained for the artifact data, exemplified by the umbrella–lampshade case. Modifying the appearance while retaining the function had a greater effect on similarity than on categorization, but the likelihood of categorization still fell from 9.9 to 7.0 on the 10-point scale. Likewise, the object that retained the correct appearance but was intended and used for the wrong function managed 3.6 on the scale. Subjects were far from being confident about the irrelevance of mere appearance in making category judgments about artifact kinds.

There is other research which supports the findings reported here. Studies by Malt (Malt, 1994; Malt & Johnson, 1992) have found evidence against the common assumption that natural kind concepts have cores defined by essences and artifacts have cores defined by function. Malt (1994) showed that people's use and understanding of the concept WATER could not be simply related to the percentage of H_2O contained in the fluid. Factors such as use by humans appeared to influence the cate-

gorization of fluids as kinds of water. Malt and Johnson (1992) considered the role of functional use in defining artifacts, and found that both physical appearance and intended function had an effect on the categorization of hypothetical objects.

It appears to be an unavoidable conclusion that naive subjects refuse to share the intuitions of the experts when it comes to concept definitions. Philosophers and psychologists may agree that natural kinds like WATER or LEMON should be defined in terms of some essential and hidden structural property. The subjects in these experiments and in Malt's studies do not apparently share this view. The question then is whether the psychology of concepts should be modelling concepts per se, or people's categorization behavior. This question has been the subject of a continuing debate between philosophers and psychologists (Rey, 1983, 1985; Smith, Medin, & Rips, 1984), and remains a live issue (Hampton, 1994; Sutcliffe, 1993). The position taken by the present author is that the most central issue in the psychology of concepts is people's behavior in categorization tasks, and not the question of whether the beliefs that drive people's categorization are correct.

If the support for the Binary Model is so equivocal, why does the model have such a strong intuitive appeal? The answer may be that for certain concepts, and given a certain level of education, people may learn to maximize the weights of some features relative to the others to the point where a DF/CF distinction becomes apparent. The findings reported here suggest that this state of affairs may be less common than is supposed. It is almost certainly a function of human civilisation (as embodied for example in science, philosophy, law making and civil administration) to take "folk" concepts and to render them more clearly defined. There will be a strong temptation for those with a great deal of education (such as researchers and thinkers concerned with concepts) to suppose that the degree of conceptual clarity that they have achieved themselves is also common to others. However this remains an empirical question and the failed attempt to find clearly defined concepts in the experiments reported here is evidence against that supposition.

While providing no support for the Binary Model, the results reported here are consistent with the radical essentialism proposed by Medin and Ortony (1989). If categorization is driven by essences which are hidden from simple inspection, and may even be unknown to many subjects, then one would expect subjects to be uncertain about the categorization of the unusual examples described in the experiments. A zebra is a zebra because of its zebra-like essence—whatever that may be. Individuals with little or no schooling in cell biology and genetics, and little practical experience of animal husbandry, may be quite willing to believe that this hidden essence can be altered through such environmental accidents as a special diet during development. Paradoxically, the theory of psychological essentialism predicts a much greater degree of uncertainty in categorization than the Classical Model. If the person categorizing some instance (a) does not know what the essence of the concept consists of and/or (b) does not know if the instance being categorized possesses that essence, then categorization is likely to be probabilistic and to be inferred on the basis of the available information—which itself may well simply reflect similarity of the instance to a prototype representation of the concept. Probabilistic categorization based on similarity (as seen in the present results) may therefore also reflect concept representations with unknown essences.

Feature Independence

The second test of Prototype Theory concerned the question of how feature matches are combined to derive a measure of similarity to the prototype. The results of the analysis of Feature Independence showed that, as indexed by categorization

probability, the effect of features on similarity was greater the higher the initial similarity. Manipulation of the presence or absence of the same feature set produced stronger effects for concept exemplars which were high probability category members than for those that were marginal to the category. This result is inconsistent with the predictions of traditional versions of Prototype Theory, which would expect stronger effects of feature change on instances that are closer to the category border (as defined by the 50% categorization probability threshold). It is also inconsistent with any similarity based theory of categorization in which the computation of similarity assumes Feature Independence (as in Tversky, 1977). Prototype Theory is therefore in need of revision.

It may at first appear to be paradoxical that the effect of removing a set of superficial characteristic features is greatest when an instance is clearly in the category. It could be expected that ceiling effects on categorization probability would be bound to reduce the size of the effect of removing the features when the instance is otherwise a clear category member. If, for example, one considered a chair that was perfectly prototypical except that it had six legs rather than four, then one would not expect to see the categorization probability drop by much. The analysis presented here however was not concerned with this region of the similarity curve. By only manipulating *sets* of characteristic features together, the design of the present experiments combined their individual effects, so that in most of the concepts there was a significant drop in categorization probability when the set of CF were removed from the DF+CF+ scenario. In addition, the analysis of data shown in Fig. 2 eliminated concepts where the CF effect was very weak or not present, or where average categorization probability for DF+, with and without the CF present, was close to 1. It is not argued from the present data that there would be no region on the curve at which categorization probability would be at asymptote (as in the region to the right of Fig. 1). However, because of the difficulty of finding CF which did not affect categorization, this region of the curve was not well sampled in the present experiments. Further experiments would need to be run, sampling from a wider range of scenarios within each concept, in order to provide a clearer picture of the full relation between similarity (feature match) and categorization probability.

The pattern of data found here is in fact consistent with other results in the literature on similarity and categorization. First, Medin and Shaffer (1978) (see also Nosofsky, 1988), argued that similarity between two exemplars of a category should be assessed not as a sum of matching features, but as a *product* of matching features. According to their algorithm for assessing similarity, the effect of any single feature would then be multiplied by the presence of other positive features, producing just the pattern found in the present experiments. They saw this aspect of their model as being a way of incorporating *necessary* features, since if any feature dimension were to produce a match value of zero, then the similarity of the two exemplars would by definition also be zero. There is obviously more needed to this formulation, for what is true for a defining feature would also be true for a characteristic feature. If a bird's ability to fly is just so bad that there is no match at all to the "flying" feature (penguins and ostriches come to mind as good examples), then of course the similarity would also reduce to zero, and penguins and ostriches would not be birds. In order to differentiate between necessary and nonnecessary features with a multiplicative rule it is necessary to place some arbitrary constraint on whether feature matches can fall to zero or only to some nonzero positive value. Only necessary features would be allowed to have the former property. The degree to which the value of a nonmatching feature falls below 1 could then be used as a pa-

rameter of the weight attached to a feature. In effect each feature would be given a weight between 0 and 1 corresponding to the proportional decrease in similarity resulting from negating that feature from an instance. A weight of 0 would indicate a necessary feature (similarity reduced to 0% of its previous level), while a weight of 1 would indicate an irrelevant feature (similarity remaining at 100% of its previous level.

Such a scheme would still be a Prototype (as opposed to a Binary) Model, provided that the threshold criterion placed on similarity for category membership was greater than zero. With a criterion greater than zero, an instance could in principle possess all the necessary features, but still not belong in the category (because it lacked too many of the nonnecessary features). With a threshold set at zero, the model would reduce to the Binary Model, since the only way for an instance to be rejected would be for it to lack a necessary feature, that is for it to have at least one feature with a weight of zero. It may be argued that as often formulated, Rosch's Prototype Theory does not admit of *any* necessary features. Thus the prototype is said to consist of a set of features "none of which are necessary for membership." However on reflection it is clear that there must always be a set of domain general features which are common to the whole class, and hence necessary. Thus all FRUITS are organic, all types of FURNITURE have a human function, all FISH are living (or were once living). Such necessary features are commonly found in prototype concepts (Hampton, 1979), and do not undermine the central point of a prototype structure, which is that taken in conjunction such features are *not sufficient* for membership, and so do not differentiate a category from its immediate contrasting sets.

A second source of confirmation for the present results comes from Shepard's theory of stimulus generalization (Shepard, 1987) which proposed that similarity is an exponentially decaying function of distance in feature space. The positively accelerated curve for the data in Fig. 2 is consistent with an exponential function relating degree of feature match to underlying similarity. Prototype Theory could perhaps be rendered compatible with the present results by the interpolation of an exponential transformation relating the sum of weighted features to the underlying similarity dimension. This proposal is actually effectively equivalent to the previous suggestion of a multiplicative combination of feature matches since:

$$\text{Similarity} = e^{(f_1 + f_2)} = e^{(f_1)} \cdot e^{(f_2)} \qquad (5)$$

Finally, a study by Tversky and Gati (1982) also provides an interesting parallel to the result reported here. In one experiment, Tversky and Gati compared the pairwise similarity of four stimuli which varied on two dimensions. They showed that the rated similarity of a pair of stimuli which matched exactly on the first dimension, and differed by (say) 10 units on the second dimension was greater than the similarity of a stimulus pair which differed by five units on both dimensions. Thus the positive effect on similarity of one exact matching dimension was greater than that of two partially matching dimensions.

Mathematical modeling of concept structure has previously been largely confined to artificially constructed stimuli (for example, Nosofsky, 1988). A problem with applying a more quantitative approach to natural concepts has been an overdependence on typicality ratings as the critical dependent variable. Typicality is an important phenomenon in its own right, but it is clear that there are many influences on rated typicality, over and beyond the degree to which an instance fits the concept representation (Barsalou, 1985), to the point where typicality cannot be taken as a direct measure of degree of category membership (see also Rips, 1989). It is to be hoped that the systematic manipulation of category features, together with the use of categorization probability as a dependent

measure may provide a more robust measure of categorization behavior which can be used to develop more specific models of conceptual structure.

Appendix A

Examples of the Six Scenarios for Two of the Concepts from Experiment 3

(1) Concept: UMBRELLA

DF+CF+ Waterproofed cloth stretched over a wire frame to form a dome shape with a long handle. It was designed to be carried over one's head, in order to keep off the rain. It can be collapsed when not in use.

DF(?)CF+ Waterproofed cloth stretched over a wire frame to form a dome shape with a long handle. It was designed as a way of protecting people from acorns and small twigs falling on them when seated under oak trees in the park. It can be collapsed when not in use.

DF−CF+ Waterproofed cloth stretched over a wire frame to form a dome shape with a long handle. It was designed as a kind of reflector for indoors TV reception. It can be collapsed when not in use.

DF+CF− Waterproof plastic tacked to a light wooden frame in the shape of a hexagon with a short handle. It was designed to be carried over one's head, in order to keep off the rain. It can be collapsed when not in use.

DF(?)CF− Waterproof plastic tacked to a light wooden frame in the shape of a hexagon with a short handle. It was designed as a way of protecting people from acorns and small twigs falling on them when seated under oak trees in the park. It can be collapsed when not in use.

DF−CF− Waterproof plastic tacked to a light wooden frame in the shape of a hexagon with a short handle. It was designed as a kind of reflector for indoors TV reception. It can be collapsed when not in use.

(2) Concept: UNCLE

(Common introduction:)
Sam is a little boy aged 5 who lives at home with his parents. Sophie (aged 22) is his mother. Is this person really Sam's uncle?

DF+CF+ A cheerful man aged 35 who comes to the house at Christmas bringing presents for Sam. He is a regular visitor to the house. His sister is Sam's mother, Sophie.

DF(?)CF+ A cheerful man aged 35 who comes to the house at Christmas bringing presents for Sam. He is a regular visitor to the house. When he was 1 year old, his parents split up, and his father married Sophie's mother, so he is Sophie's step-brother.

DF−CF+ A cheerful man aged 35 who comes to the house at Christmas bringing presents for Sam. He is a regular visitor to the house. He is a family friend of Sam's mother, Sophie.

DF+CF− A young boy aged 7 who Sam has never met, and who lives in Australia. His sister is Sam's mother Sophie.

DF(?)CF− A young boy aged 7 who Sam has never met, and who lives in Australia. When he was 1 year old, his parents split up, and his father married Sophie's mother, so he is Sophie's step-brother.

DF − CF − A young boy aged 7 who Sam has never met, and who lives in Australia. He is a family friend of Sam's mother, Sophie.

APPENDIX B

Worked-Out Example of the Data Transformations Used to Derive Three Pairs of Data Points for Fig. 2 from the Set of Six Scenario Frequencies Produced for a Concept (Frequencies are Illustrative)

Scenario	Frequency *f*	Modified *f*	Probability	*z* value
DF+ CF+	12	11.75	.979	+2.04
DF(?) CF+	8	8	.667	+0.43
DF− CF+	3	3	.250	−0.67
DF+ CF−	10	10	.833	+0.97
DF(?) CF−	5	5	.417	−0.21
DF− CF−	2	2	.167	− .097

Defining feature	CF effect	Standardized CF effect
DF+	1.07	1.597
DF(?)	0.64	0.955
DF−	0.30	0.448
Mean	0.67	1.000

Note. CF effect = (Difference in *z* values for CF+ and CF−). eg., 2.04–0.97 = 1.07. Standardized CF effect = CF effect/average of the three CF effects, eg., 1.597 = 1.07/0.67.

REFERENCES

ARMSTRONG, S. L., GLEITMAN, L. R., & GLEITMAN, H. (1983). What some concepts might not be. *Cognition*, **13**, 263–308.

BARSALOU, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **11**, 629–654.

BARSALOU, L. W. (1987). The instability of graded structure: implications for the nature of concepts. *In* U. Neisser (ed.) *Concepts and Conceptual Development: Ecological and Intellectual Bases of Categories*. Cambridge: Cambridge Univ. Press.

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

GOLDSTONE, R. (1994) Similarity, Interactive Activation, and Mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 3–28.

GOLDSTONE, R., MEDIN, D. L., & GENTNER, D. (1991). Relational similarity: The non-independence of features in similarity judgments. *Cognitive Psychology*, **23**, 222–264.

HAMPTON, J. A. (1979). Polymorphous concept in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, **18**, 441–461.

HAMPTON, J. A. (1981). An investigation into the nature of abstract concepts. *Memory and Cognition*, **12**, 151–164.

HAMPTON, J. A. (1988). Overextension of conjunctive concepts: evidence for a unitary model of concept typicality and class inclusion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **14**, 12–32.

HAMPTON, J. A. (1991). The combination of prototype concepts. *In* P. Schwanenflugel (Ed.), *The psychology of word meanings*. Hillsdale, NJ: Erlbaum.

HAMPTON, J. A. (1993). Prototype models of concept representation. In I. van Mechelen, J. A. Hampton, R. S. Michalski, & P. Theuns (Eds), *Categories and concepts: Theoretical views and inductive data analysis*. London: Academic Press.

HAMPTON, J. A. (1994). *What psychologists talk about when they talk about concepts*. Paper presented to the Society for Philosophy and Psychology, Memphis TE, May.

HAMPTON, J. A., & DUBOIS, D. (1993). Psychological models of concepts: Introduction. *In* I. van Mechelen, J. A. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis*. London: Academic Press.

KEIL, F. C. (1986). The acquisition of natural kind and artifact terms. In W. Demopoulos & A. Marras (Eds.), *Language learning and concept acquisition: Foundational issues*. Norwood, NJ: Ablex.

KEIL, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

KEIL, F. C., & BATTERMAN, N. (1984). A Characteristic-to-Defining shift in the development of word meaning. *Journal of Verbal Learning and Verbal Behavior*, **23**, 221–236.

KRIPKE, S. (1972). Naming and necessity. In D. Davidson & G. Harman (Eds.), *Semantics of natural language*. Dordrecht: Reidel.

LANDAU, B. (1982). Will the real grandmother please stand up? The psychological reality of dual meaning representation. *Journal of Psycholinguistic Research*, **11**, 47–62.

MALT, B. C. (1990). Features and beliefs in the mental representation of categories. *Journal of Memory and Language*, **29**, 289–315.

MALT, B. C. (1991). Word meaning and word use. In

P. Schwanenflugel (Ed.), *The psychology of word meanings*. Hillsdale, NJ: Erlbaum.

MALT, B. C. (1994). Water is not H_2O. *Cognitive Psychology*, **27**, 41–70.

MALT, B. C. & JOHNSON, E. C. (1992). Do artifact concepts have cores? *Journal of Memory and Language*, **31**, 195–217.

MARGOLIS, E. (1994). A reassessment of the shift from the classical theory of concepts to prototype theory. *Cognition*, **51**, 73–89.

MARKMAN, A. B., & GENTNER, D. (1993). Structural alignment during similarity comparisons. *Cognitive psychology*, **25**, 431–467.

MCCLOSKEY, M., & GLUCKSBERG, S. (1978). Natural categories: well-defined or fuzzy sets? *Memory and Cognition*, **6**, 462–472.

MEDIN, D. L., & SHAFER, M. M. (1978). Context theory of classification learning, *Psychological Review*, **85**, 207–238.

MEDIN, D. L., & ORTONY, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and Analogical Reasoning*. Cambridge: Cambridge Univ. Press.

MILLER, G. A., & JOHNSON-LAIRD, P. N. (1976). *Language and Perception*. Cambridge Mass: Harvard Univ. Press.

MURPHY, G. L. (1993). Theories and concept formation. In I. van Mechelen, J. A. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis*. London: Academic Press.

MURPHY, G. L., & MEDIN, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, **92**, 289–316.

NOSOFSKY, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **14**, 700–708.

OSHERSON, D., & SMITH, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, **9**, 35–58.

OSHERSON, D., & SMITH, E. E. (1982). Gradedness and conceptual conjunction. *Cognition*, **12**, 299–318.

PUTNAM, H. (1975). The meaning of 'meaning'. In *Mind, language and reality, volume 2: Philosophical papers*. Cambridge: Cambridge Univ. Press.

REY, G. (1983). Concepts and stereotypes. *Cognition*, **15**, 237–262.

REY, G. (1985). Concepts and conceptions: A reply to Smith, Medin & Rips. *Cognition*, **19**, 297–303.

RIPS, L. J. (1989). Similarity, typicality and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and Analogical Reasoning*. Cambridge: Cambridge Univ. Press.

ROSCH, E. (1973). On the internal structure of perceptual and semantic categories. *In* T. E. Moore (Ed.), *Cognitive development and the acquisition of language*, New York: Academic Press.

ROSCH, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, **104**, 192–232.

ROSCH, E., & MERVIS, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, **7**, 573–605.

SHEPARD, R. (1987). Toward a universal law of generalization for psychological science. *Science*, **237**, 1317–1323.

SMITH, E. E., & MEDIN, D. L. (1981). *Concepts and categories*. Cambridge, MA: Harvard Univ. Press.

SMITH, E. E., MEDIN, D. L., & RIPS, L. J. (1984). A psychological approach to concepts: Comments on Rey's "Concepts and Stereotypes." *Cognition*, **17**, 265–274.

SMITH, E. E., SHOBEN E. J. & RIPS, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, **81**, 214–241.

SMITH, E. E., & SLOMAN, S. A. (1994). Similarity-versus rule-based categorization. *Memory and Cognition*, **22**, 377–386.

SUTCLIFFE, J. P. (1993). Concept, class, and category in the tradition of Aristotle. In I. van Mechelen, J. A. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis*. London: Academic Press.

TVERSKY, A. (1977). Features of similarity. *Psychological Review*, **84**, 327–352.

TVERSKY, A., & GATI, I. (1982). Similarity, separability, and the triangle inequality. *Psychological Review*, **89**, 123–154.

(Received July 5, 1994)
(Revision received February 24, 1995)