Overextension of Conjunctive Concepts: Evidence for a Unitary Model of Concept Typicality and Class Inclusion

James A. Hampton
The City University, London, England

Four experiments investigated how people judge both the typicality and membership of items in conjunctive concepts such as *school furniture* or *sports which are games*. Judgments of membership in conjunctions were overextended, and there was asymmetry between the constituent concepts in their influence on relative conjunctive concept membership. The results are discussed in the light of recent theoretical disputes about the modeling of concept representations and the process of forming conjunctions (Cohen & Murphy, 1984; Osherson & Smith, 1981, 1982; Smith & Osherson, 1984). A theory is proposed in which constituent intensions are combined to form a composite prototype for the conjunction. Membership in both single and conjunctive concepts is then determined in the same unitary fashion, by placing a membership criterion on the perceived similarity of possible exemplars to the prototype.

An important issue in the study of natural language concepts is the way in which common semantic concepts combine to form conjunctions. For a wide range of concepts, two very reliable phenomena have been established: One, members of concept categories vary in their representativeness, and two, for many concepts the boundary around the class of concept members is unclear or "fuzzy." However, to date few empirical studies have considered what happens when two of these fuzzy concepts are placed in conjunction. Do conjunctions show similar typicality and fuzziness phenomena? If so, how can they be related to typicality in the constituent concepts? Do fuzzy concepts follow the same logical rules for conjunction as well-defined concepts? The first aim of this article is to provide empirical evidence that may begin to answer these questions.

The current interest in concept conjunctions comes largely from their relevance to the basic question of concept definitions: that is, how a concept picks out an extensional set of concept members. A second aim of this research is therefore to use the study of conceptual combination to provide important constraints on models of conceptualization. In particular, it will be argued that evidence on conjunctions can have theoretical implications for distinguishing the following two opposing accounts of concept definitions.

According to the first view (Hampton, 1979, 1981; McCloskey & Glucksberg, 1979; Rosch, 1978), variations in typicality reflect differences in similarity[1] to a summary representation of the concept (the concept *prototype*). Concept membership is determined by placing a criterion on the similarity dimension, so that members are just those items with similarity to the prototype greater than the criterion. The two phenomena are thus attributed to a single underlying factor: the *unitary hypothesis*.[2] As evidence for this view, research has shown that the probability of an item being classed in a category increases in a smooth fashion with rated typicality (McCloskey & Glucksberg, 1978), and with the number of category attributes possessed by an item (Hampton, 1979). In a regression study, Hampton (1984) found that classification probability was again predicted by item typicality with no residual effect of familiarity or category production frequency.

The alternative *binary* view states that typicality and set membership are determined in essentially different ways. According to this view, although typicality judgments depend on similarity as described earlier, set relations such as class inclusion, negation, conjunction, and disjunction follow the standard logic of sets, in which membership in a set is an all-or-none affair with no gradations. Smith, Shoben, and Rips (1974) distinguished between typicality effects and set membership in this way. Concepts were proposed to have *defining* features, which provide a necessary and sufficient determination of set membership, and *characteristic* features, which only determine typicality. For example, feathers and two legs are among the features used to determine what is a *bird*, whereas flying and sing-

[1] Similarity is used here in a general sense. It is not necessary to assume any specific similarity metric because the arguments remain on a general level.

[2] Another type of model, exemplar models, makes similar unitary assumptions, although in this case the similarity function is computed across a range of the most typical exemplars stored in memory, as opposed to a single prototype (Medin & Schaffer, 1978, Smith & Medin, 1981).

ing only distinguish typical *birds* from others. A well-known problem with this distinction (Hampton, 1979; McCloskey & Glucksberg, 1978, 1979) is that unlike the well-worn example of *birds,* many semantic categories have no easily identifiable defining features that could provide the required criteria for determining set membership (see also McNamara & Sternberg, 1983). It would also be hard to account for the number of borderline cases and the looseness of many class inclusion statements if such features really existed. For example, Hampton (1982) showed that items such as *car-seats* may belong to an immediately superordinate class *chairs,* but not belong to the superordinate of *chairs-furniture.* As well as arguing against nested defining features for concept hierarchies, this intransitivity suggests that statements such as "Chairs are furniture" are not to be taken as universally quantified statements.

Despite these reservations about defining features, the binary view has received support from three other sources. First, Osherson and Smith (1981) argued strongly for the view by pointing to a series of problems in the formalization of prototype theory using fuzzy set theory (Zadeh, 1965). From these problems they claimed there were serious if not fatal flaws in the prototype theory itself. Their arguments have received considerable attention (for details see Cohen & Murphy, 1984; Hampton, 1983; Jones, 1982; Lakoff, 1987; Osherson & Smith, 1981, 1982; Smith & Medin, 1981; Smith & Osherson, 1984; Thagard, 1983; Zadeh, 1982). Perhaps the most telling problem they raised concerned the subject of the research reported here, namely, the formation of conjunctive combinations of fuzzy sets. Briefly, they showed that if the extent to which an item belongs to a category is given a value between 0 and 1 (as proposed in fuzzy logic, Zadeh, 1965), no simple general rule could exist that would map the membership value of an item in two constituent sets into its membership value in the conjunction of those sets. Of particular embarassment for the fuzzy logic position (which proposed either a minimum or a multiplicative function for conjunction) was the existence of items that were better examples of a conjunction than they were of either constituent concept (for example, *guppy* as a *fish,* as a *pet,* or as a *pet fish*). They concluded from their analysis of the prototype view that fuzzy set notions are possibly appropriate only to judgments of typicality gradients, whereas class inclusion, conjunction, disjunction, and negation follow traditional set logic.

A second source of support for the binary view was Armstrong, Gleitman, and Gleitman's (1983) demonstration that typicality judgments could consistently be given to concepts that are presumably not fuzzy at all, for example odd and even numbers or male and female. This demonstration of typicality variation in apparently well-defined concepts threw doubt on the supposed common mechanism underlying typicality and set membership decisions by dissociating the two phenomena (see also Bourne, 1982). At least some concepts, it appears, can have typicality gradients without having prototype definitions of membership.

A third line of argument for the binary view concerns "natural kind" concepts such as *tiger, lemon,* or *gold.* In such concepts, typicality variations may reflect how well an instance fits some kind of heuristic *recognition procedure* for set members, whereas true category membership may rely on less accessible

criteria that are possibly available only to experts (see Keil, 1986; Miller & Johnson-Laird, 1976; Putnam, 1975).

Although these arguments do little to disprove the idea that *some* concepts (or even many) have prototype definitions, more evidence is needed to confront this issue. The study of conjunctive concepts can provide such evidence. Abandonment of the traditional *common element* form of definition leads the unitary hypothesis to predict inconsistencies in the interpretation of fuzzy set conjunction. In particular it can predict that when an item is a good member of one concept and is not quite a member of a second, such an item may still be judged to belong to their conjunction. Conjunctions should suffer overextension as a result of compensation. The prediction of overextension follows naturally from the unitary view, given Osherson and Smith's (1981) demonstration that an item's similarity to a conjunction can sometimes be higher than its similarity to a constituent of that conjunction (the *guppy effect*). If probability of class membership is also a function of similarity, we can predict that at some point on the scale, items may be found that also have greater *probability of membership* in the conjunction than in the constituent. The prediction is apparently counter intuitive. It expects people to agree with two inconsistent statements: "X is an A that is a B," but also "X is not an A." Thus, for example, *blackboard* may be considered to be *school furniture,* but not to be *furniture* (although the subject agrees that *school furniture* is a type of *furniture*), or *chess* may be a *game* which is a *sport,* yet not be a *sport* simpliciter. If such logically "inconsistent" views can be demonstrated for a particular set of concepts, showing that concept conjunction in those cases does not follow standard set logic, then such a result would lend support to the unitary hypothesis and would be at odds with the binary view as currently stated. On the other hand, if people reserve membership in the conjunction solely for those items judged to belong to both constituent categories, regardless of typicality, then the binary view will be vindicated. Categorization and judgments of typicality would rely on different semantic information, a serious blow to prototype theory.

Four experiments will be described. Experiment 1 used simple noun–noun and adjective–noun compounds, such as *school furniture,* or *protective clothing,* and compared membership in the compound category with membership in the "head" noun category (*furniture,* or *clothing*). Using noun plus relative clause constructions to express the conjunction, the remaining three studies undertook a more quantitative examination of conjunction formation. Experiments 2 and 3 examined the specific conjunction "sports which are also games" and varied item membership in each category in a continuous and systematic way. Experiment 4 provided generality by using six new conjunctions of concepts.

Experiment 1

The first experiment took as examples of conjunctions noun–noun compounds like *office furniture* and *sports vehicles.* Linguistic analysis (Cohen & Murphy, 1984; Levi, 1978) has stressed the flexibility with which the meaning of a compound is related to the meanings of its constituent parts. For example, in each of the following cases—*a bicycle repair, an expert repair, a Scotch-tape repair*—the first noun (the qualifier) fills a differ-

ent semantic role. However, it is generally assumed that the head noun (the final noun in a compound) does in fact define the general category in which the concept lies.[3] Thus, for most compounds, it is fair to assume that the phrase "a BA" can be paraphrased as an explicit conjunction "an A that is qualified in some respect by B." For example, the above cases are repairs which are *done to* bicycles, *by* experts and *with* Scotch-tape, respectively (given a normal context of utterance). Context may readily change the semantic role of the qualifier (e.g., the terms could refer to repairs that *require use of* a bicycle or repairs *made for* experts), and additional information may be implied by the qualifier (for example, emergency repairs are usually temporary), but all the examples still remain firmly in the head-noun class of *repairs*.

If the unitary hypothesis applies to such concepts, this general assumption may turn out to be false, even when an explicit noun-plus-relative-clause paraphrase rules out idiomatic or metaphorical usage. The unitary hypothesis predicts that items that fit the qualifier extremely well (in its intended indirect sense) should be included in the conjunction, even when they are not generally considered to belong to the head-noun category. The binary hypothesis, on the other hand, predicts a strict class inclusion to hold between the conjunction and the head-noun category.

Method

Design and materials. Eight noun compounds were formed from three general categories. All eight could be analyzed as an explicit conjunction by a noun-plus-relative-clause phrase (e.g., *office furniture* is *"furniture* found and used in *offices"*) The eight compounds were *office furniture, garden furniture, school furniture, kitchen furniture* and *church furniture; sports vehicles and aquatic vehicles;* and *protective clothing.* Between 8 and 18 specific instances for each compound were selected, and each subject rated the membership of each item (a) in the subset defined by the compound itself and (b) in the general category (*furniture, vehicles,* or *clothing*). Half of the subjects rated the compound first, and half rated the category first. The items included both positive and negative examples for both the compound and the category, as well as items intuitively borderline to the concepts. There were 127 instances in all.

Procedure. Subjects completed a booklet in their own time and were paid on returning it. The instructions asked subjects to decide first if each item was a member of the set named at the head of the list. If it was a member, they chose an appropriate positive number from 1 to 3 to indicate degree of typicality. If it was *not,* they chose a negative number from −1 to −3 to indicate relatedness as a nonmember. The response scale and its interpretation were typed at the top of each sheet. A zero response was allowed if a subject was unable to decide on a classification. Thus, each item received a number on a scale between +3 (very typical) and −3 (unrelated), on which the category border was explicitly defined as zero.

On returning the booklet, the subject was given an unexpected final sheet to complete. There were two sections. In the first, subjects were given eight sentences of the type, "Garden furniture is a type of furniture," one for each compound, and were asked to judge the statements as true or false. These were a check on whether the compounds were perceived as belonging to the head-noun class. In the second section subjects were shown eight pairs of sentences of the type, "A. All garden furniture is also furniture. B. Some garden furniture is not furniture." and decided which of each pair was more true. The purpose of this second section was to see whether subjects may be aware of counterex-

amples to the universally quantified class inclusion "All X are Y." If the explanation of category intransitivity offered by Hampton (1982) is correct, namely, that categorizations are intended to be taken as generally but not universally true, then a substantial number of subjects may be found to agree that "X is a type of Y," at the same time as agreeing that "some X is not Y." Such a result would be a striking demonstration that categorizations need not imply extensional class inclusion.

Subjects. A total of 22 students at The City University in London acted as paid volunteer subjects. All were unaware of the purpose of the experiment.

Results

Each subject's pair of ratings of any particular item could take one of four patterns on the basis of the sign of each rating. Taking pairs in the order [compound, general category], the critical pattern is [+−], where an item belongs to a conjunction, say *protective clothing,* but not to its category *clothing.* For the following analysis, zero responses (less than 4%) were treated conservatively, being counted as negative for a compound, and positive for a general category. Ten responses of 00 were excluded.

The overall mean percentage of [+−] responses was 26 ± 7% for subjects rating the compound first, and 22 ± 8% for those rating the category first (95% confidence limits are quoted). The responses were not randomly distributed across items. The correlation across items within each subset of [+−] frequency between the two subject groups averaged 0.602, and was significant in all but three of the eight subsets. For *school furniture* the correlation was almost significant, but for the two subsets of *vehicles* it was nearly zero. For the remaining compounds there were therefore particular items reliably producing inconsistent responses. Appendix A lists examples with at least 10 [+−] responses ($n = 22$), with their response distribution, and the two mean ratings. (For *vehicles,* items with at least 5 [+−] responses are shown.) The mean ratings confirmed the overextension of the compounds. The large majority of items in Appendix A had positive conjunction ratings and negative general category ratings.

The probability of positive categorization in the conjunction was compared with that for the general category for all 127 items. Although 93 items were more likely to be in the conjunctive compound than in the general category, only 22 items showed the reverse effect. This pattern was unaffected by the order of making the two ratings.

On the final questions, only 6% of the subjects on average failed to agree with the "XY is a type of Y" sentences, which confirmed that there was no obvious metaphoricity in the Y term (as might be the case for example with "A sitting duck is a kind of duck" or "A clothes horse is a kind of horse"). However, on average 66% of the subjects rejected the universally quantified sentences in favor of the "some XY are not Y" sentences. Those subjects agreeing with universal quantification showed significantly less inconsistency than those rejecting it,

---

[3] Two exceptions are idiomatic metaphors such as *sitting duck* or *hobby horse,* and another class of examples such as *a counterfeit dollar, a decoy duck,* or *a false economy,* where the qualifier serves to indicate that the example belongs in the category in appearance only and *not* in reality.

which indicated some awareness of the existence of counterexamples, and an ability to make extensional judgments, when required. Compounds with less inconsistency were also significantly more likely to be judged as conforming to "all X are Y" ($r = -0.85, p < .01, n = 8$, between percentage of inconsistency and percentage of universal quantification).

Discussion

A large number of items in this experiment were more often judged to belong in a conjunction such as *school furniture* or *protective clothing* than in the categories from which these concepts are supposedly drawn, namely *furniture* and *clothing*. Thus, items belonged in a conjunction that did not belong in one of its constituent concepts. This result is a direct extension of the intransitivity phenomenon reported by Hampton (1982). In both cases a "subset" of a category was shown to contain instances that did not belong to that category. The result makes it unlikely that subjects apply classical extensional logic to decide the reference of these compounds, although they appear to be simple conjunctions. The results also show that just as items may be more *typical* of a conjunction than of a constituent, (Osherson & Smith, 1981), so they may also be *more likely to belong*. There is therefore no need here to propose different bases for membership judgments and for typicality ratings.

An interesting parallel can be drawn with the recent demonstration by Tversky and Kahneman (1983) that the perceived likelihood of a conjunction of two events may exceed that of just one of those events. For example, given that Linda supports the Equal Rights Amendment, the chances of her being a feminist bank teller are judged as being greater than the chances of her being simply a bank teller. They argue that people base these judgments on intensional similarity or *representativeness* (Kahneman & Tversky, 1972; Tversky, 1977). It will be argued later that a very similar process is involved in deciding conjunctive category membership.

Given that people are not apparently treating compounds as proper subsets, how might we explain the interpretation of compound concepts? One source of explanations[4] lies in the semantics of noun–noun compounds. Cohen and Murphy (1984) noted that the meaning of a compound is rarely a simple conjunction, but rather involves the *modification* of the meaning of one word by the meaning of the other. (Smith, Osherson, Rips, Albert, & Keane, 1985, make a similar proposal.) For example, *school furniture* is not the class of things that are both *schools* and *furniture*. Compounds generally invoke an implicit ground mediating the combination of the concepts, for instance, "furniture *found and used in* schools." If taken as just arguing for the need to expand the qualifier noun into an explicit class before identifying the conjunction, this view would not of course explain the present results because it would not explain how the head-noun class came to be overextended. *School furniture* was not that subset of the class *furniture* that is found and used in schools; it included objects not normally in the general class of *furniture*. Cohen and Murphy's argument therefore has to be taken further to suggest that the concept *school* modifies the concept of *furniture* in a more radical sense, perhaps by providing a new context in which the word receives a different sense (Murphy & Medin, 1985). Unfortunately, it is difficult to define

this process. The model to be described in the final section suggests some mechanisms for interactive concept modification.

A second account of the results might be offered in terms of metaphorical extension. Perhaps category concepts become overextended in compounds similarly to the way in which words can be used to refer metaphorically to things outside their extension. In this case, subjects could be applying logical conjunction to an extended head-noun class. There are two counters to this suggestion. First, metaphorical extension generally involves applying a term to some semantic domain which is far removed from the normal reference of the term. (Consider, for example, India as the jewel in Queen Victoria's crown.) In the present case, however, the overextension is to other objects in the same domain which happen to fall just outside the normal reference of the category term. Second, it seems likely that if asked of a true metaphor if it was true that "X is a type of Y," where X was the metaphorical and Y the literal sense of the term, a large number of people would say no. (Is fishing for compliments a type of fishing?) However, on average 94% of the present subjects rated these statements as true. Further empirical data is needed here to know whether this type of question would distinguish clear cases of metaphor from the kinds of overextension found in this experiment. It is also possible that metaphor is just one of a number of ways in which classes become extended, and that compound formation is another. However, this position would hardly count as an explanation of the data.

A third account of the overextension involves metonymy: the use of an exemplar or subset term to refer to a more general category (see Lakoff, 1987). For instance, *furniture* may have a general sense of "useful artifacts with particular functions and appearance" and a more specific sense of *domestic furniture*, including the specification of a domestic setting and domestic functions. Following the narrow sense of the term, a subject may judge that a *blackboard* is not *furniture*, but then when forced by the compound to take the more general sense, he may decide that it is. This account would be supported if subjects rating *school furniture* first had been less likely to overextend the conjunction (being primed to the general sense of the term). However, there was no such effect in the data.

A final possibility is that the compound terms in some way encourage a more "technical" sense of the head-noun term, by changing the perceived linguistic context of usage of the term. It is not clear in what sense undergraduate students may be said to possess technical concepts of terms like *furniture* or *clothing*—such domains being untechnical—but one could imagine that the weight attached to particular kinds of attribute may be increased, given a particular context of usage.

The most plausible of these accounts is probably the modification account (Cohen & Murphy, 1984). Noun–noun compounds differ in important ways from the sum of their parts—for instance, as typified by a relative clause paraphrase—because of the interaction of the intensional information of each concept. As a result, apparently conjunctive concepts are overextended in a way that is consistent with the unitary hypothesis

---

[4] Because the remaining experiments do not use compounds, some discussion of this question is worthwhile here. More general consideration of conjunction formation will be presented in the General Discussion.

but contradicts the idea of a binary separation of class membership and typicality decisions for these compounds. This view will be taken up in some detail in the model presented in the General Discussion.

The following three experiments sought to broaden the scope of the study while avoiding some of the problems associated with compounds, by considering cases where two noun categories of equivalent generality are placed in an explicit conjunction. If overextension is still found for a concept such as "A which are B," where A and B are both familiar noun classes, then the range of examples for which people do not use set logic for forming conjunctions would be widened, thus increasing the generality of the unitary hypothesis, while further limiting that of the binary view.

Experiment 2

Experiments 2 and 3 took a single pair of categories and obtained three sets of category membership ratings: A, B, and their conjunction A&B. The aim was to test the generality of the overextension observed in Experiment 1. It was also to provide a more quantitative description of the degree of overextension in such conjunctions. Another aim was to compare membership in the conjunction A&B with that in its converse B&A, and so to test the commutativity of the conjunctions. Note that neither classical set theory nor for that matter any other current theory of concept conjunction (intensional or extensional) would predict a priori any difference between the two forms.

Method

Design and materials. The two categories chosen for this study were *sports* and *games*. Apart from acknowledging the influence of Wittgenstein (1953) in this choice of domain, the two concepts are well suited for this study, in being at a similar level of generality and partially overlapping. In the domain of recreational activities there are examples that either fall in both categories, in just one or the other, or in neither. "Sportness" and "gameness" can therefore be independently varied across a range of activities, and the corresponding level of belonging in the conjunctions "*sports* which are *games*," and "*games* which are *sports*" measured.

The data were collected in two stages. In Stage 1, a total of 55 items were rated for sportness and gameness by using the same rating scale as for Experiment 1. Items were chosen from all four possible combinations of membership and nonmembership in the two categories. After Stage 1, a total of 12 items were omitted because they were either unfamiliar, ambiguous, or very close in meaning to other items. In Stage 2, four weeks later, the original subjects rated the 43 items once more. Half rated them as "sports which are games," and half as "games which are sports," thus, the two independent factors were the order of rating sportness and gameness at Stage 1 and the version of the conjunction given in Stage 2. Subjects were randomly allocated in equal numbers to one of four groups in a 2 × 2 design.

Subjects. The 36 volunteer subjects (both male and female) were students in London between the ages of 18 and 33 years. One subject was excluded because he clearly failed to follow the instructions for the Stage 2 task.

Procedure. The rating procedure for the two stages was identical. Items were presented in a different randomly ordered list for each rating task. Written instructions were as follows:

On the following pages you are asked to make a series of judgments

Table 1

Observed and Predicted Frequencies (F) and Percentage of Inconsistency (%I) for Each Type of Response Triple (Sport, Game, Conjunction): Experiments 2 and 3

| Response triple | Experiment 2 | | | | Experiment 3 | | | |
| | Observed | | Predicted | | Observed | | Predicted | |
	F	%I	F	%I	F	%I	F	%I
[+++][a]	527	11	526	9	593	14	597	14
[++−][b]	64		54		96		97	
[+−−][a]	156	54	253	27	217	50	366	14
[+−+][b]	183		95		216		62	
[−+−][a]	241	25	288	15	330	29	393	14
[−++][b]	79		42		132		64	
[−−−][a]	176	12	157	3	248	5	261	2
[−−+][b]	24		5		13		5	

[a] Consistent. [b] Inconsistent.

about the everyday common usage of words. In each case the question to be answered is whether or not a general category name can be applied to a particular example. For each example first decide whether you would answer 'Yes' or 'No,' and then select one of the corresponding positive or negative values to indicate the strength of your choice. If you are unable to decide, use the value zero, but avoid using this as much as possible. If you are unfamiliar with any of the examples, cross them out.

The instructions included a worked example with the category *fruit*.

In order to stress the importance of the category borderline, the instructions emphasized that subjects should decide category membership first and then reflect this decision in the sign of the scale response chosen. In Stage 1, half of the subjects received the sports list first, and the other half received the games list first. In Stage 2 these two groups were evenly divided again into those rating each version of the conjunction. Subjects wrote their names on the response booklets so that Stage 1 and 2 responses could be collated.

Results and Discussion

Four analyses of the data are presented. The first considered the consistency of the conjunction membership decisions. The second used a regression model to predict mean item rating in the conjunction from mean constituent ratings. The third analysis tabulated mean conjunction ratings for each possible combination of constituent rating responses. Finally, the commutativity of the conjunctions was assessed.

Inconsistency and overextension. The three ratings given by each subject to any particular item form a triple of [sports, games, and conjunction]. Classical set conjunction requires that c should be positive when both s and g are positive, and otherwise negative. Thus, considering simply the sign of the rating responses, [+++], [+−−], [−+−], and [−−−] represent consistent triples, [++−] indicates *underextension* of the conjunction, and [+−+], [−++], and [−−+] all indicate *overextension*. Overall frequencies of each type of triple across all subjects and items are shown in the first column of Table 1. As can be seen there was relatively little inconsistency (about 10%) for

Table 2

Means and Standard Deviations for Number of Overextensions of Two Types as a Function of Condition: Experiment 2

| | | Response triple | | | | |
| | | [+ − +] | | [− + +] | | |
Stage 1 order	Conjunction	M	SD	M	SD	N
Sport–game	Sports which are games	6.33	7.9	1.89	2.1	9
Sport–game	Games which are sports	6.78	3.8	1.11	0.6	9
Game–sport	Sports which are games	3.38	2.7	2.38	1.2	8
Game–sport	Games which are sports	4.33	2.7	3.22	2.4	9
Total		5.26		2.14		35

items in both categories and for those in neither category.[5] For items in just one of the categories, however, there was considerable inconsistency: 25% of games that were not sports, and 54% of sports that were not games were still given positive ratings for the conjunction. The overextension found in Experiment 1 was replicated.

An analysis of variance tested the effects of three factors on the number of overextensions [+ − +] and [− + +] made by each subject. Between-groups factors were the order of rating sports and games at Stage 1, and the version of the conjunction at Stage 2, and the repeated measures factor was the type of overextension ([+ − +] vs. [− + +] triples). Table 2 shows the mean and standard deviation for number of overextensions. There was a significant main effect of type of triple, $F(1, 31) = 16.37$, $p < .001$, with means of 5.26 [+ − +] triples versus 2.14 [− + +] triples. This main effect interacted significantly with Stage 1 order of rating, $F(1, 31) = 7.01$, $p < .05$. No other effects were significant ($F < 1$ in every case).

The effect of type may simply reflect the arbitrary selection of items for the list. The interaction with order shows that a constituent concept was given fewer positive ratings when rated second at Stage 1 than when rated first. For example, [+ − +] would be more likely when games were rated second than when sports were. This effect may be termed a contrast effect because the subjects tended to treat the two concepts as mutually exclusive sets, rather than overlapping. Such a tendency has been well documented in children learning word meanings (Clark, 1983; Markman, 1984).

In order to judge the true extent of overextension, two extraneous factors had to be discounted. First, the contrast effect yields apparent overextension because of the underextension of the constituent category rated second. Second, a certain degree of overextension would occur simply because of the uncertain nature of the decisions being made. People are prone to change their minds about category decisions (Barsalou, 1984; Bellezza, 1984; McCloskey & Glucksberg, 1978), and such intrasubject variability would also lead to inconsistency. To rule out these two factors, a stochastic model was constructed to predict the distribution of responses across the eight types of triple, for each item separately. The model predicted the frequency of positive conjunctive ratings assuming that subjects applied a classi-

cal conjunction rule at Stage 2 to probabilistically determined constituent memberships. The contrast effect was discounted by basing the prediction for Stage 2 on unbiased estimates of the probability of positive ratings for each constituent, on the basis of those subjects rating each category first at Stage 1. The factor of intrasubject variability was incorporated by a parameter u which was the probability that a subject "recalls" each original constituent membership decision at Stage 2. Decisions that were not recalled had to be made again, with the same probability of a positive response as at Stage 1.[6]

With the parameter u set to zero—allowing the maximum intrasubject variability—the degree of overextension predicted was 42% for [+ − +] and 24% for [− + +]. However, the amount of predicted underextension [+ + −] was twice the observed level, indicating that variability was seriously overestimated. When u was estimated for each item in order to fit the amount of underextension (assuming that underextension provides a valid estimate of subject variability), then the predicted frequency of overextensions fell well below the observed level (see Table 1). (Mean u was .512, which agreed well with a similar parameter estimated from a reanalysis of the subject variability in McCloskey and Glucksberg's, 1978, data.) It was concluded that the overextension observed could not be accounted for by the contrast effect and subject variability alone.

Appendix B shows those items for which overextension was most marked, together with observed and expected frequencies of [+ − +] or [− + +]. Also shown are the unbiased estimates of the probability of belonging to sports and games and to the conjunction. Expected frequencies were all much less than those observed, with two exceptions—wrestling and gymnastics—and probability of belonging to the conjunction was often considerably greater than the lower of the two constituent probabil-

---

[5] Even 10% inconsistency may be considered to show a failure of the classical view of concepts as well-defined sets (see McCloskey & Glucksberg, 1978).

[6] The notion of subjects "recalling" their earlier decision should not be taken too literally. As well as actual recall, the parameter u is also estimating the effects of individual differences in concept definitions and any other factors which may reduce the independence of the subject's ratings at each stage. In more detail, the model used a set of equations to predict the frequency of each type of triple. For instance, for subjects rating sports first, the equation for $p(+ − +)$ was

$$s_1(1 − g_2) \cdot [u^2 \cdot (0) + u \cdot (1 − u) \cdot g_1 + (1 − u) \cdot u \cdot (0) + (1 − u)^2 \cdot s_1 \cdot g_1],$$

(a)   (b)   (c)   (d)   (e)

where s_1 and g_1 were probabilities of being rated as sports or games when rated first, and g_2 was the probability of being a game when rated second. The labeled terms in the function are (a) probability of sports (+) and games (−) at Stage 1, assuming independence between the two ratings, which was justified by the data; (b) probability (u^2) of recalling both Stage 1 responses, and hence zero chance of a positive conjunction rating; (c) probability of recalling just the sport response times probability of now making a positive games decision; (d) if just the game rating is recalled, then the conjunction cannot be positive; and (e) if neither is recalled, the probability is the product of the two unbiased constituent probabilities.

The parameter u was estimated from the two observed frequencies for [+ + +] and [+ + −], and constrained to lie between 0 and 1. If it could not be estimated it was set to zero.

ities, generally lying between the two. Where items were good members of one set (.94 or better) and relatively marginal to the other (.18 to .88), subjects allowed good membership in the first to compensate for poor membership in the other. For example, chess is a very typical game, so that although 61% were unwilling to call it a sport, 51% still rated it as in the conjunction.

Mean item ratings. Means were calculated of the ratings given to each item for each concept. As with the probabilities of inclusion, means for the conjunction generally lay between the two constituent means. Unlike the guppy example, no case was found where the mean conjunctive rating exceeded both constituent mean ratings. Multiple regression analysis was applied to predicting mean conjunctive ratings from the mean constituent ratings.[7] Two separate equations were derived, one for each version of the conjunction. The predictor variables were the mean sport and game ratings for each item, based on the ratings given in initial rating position only. The fit of the equations was good (R = .952 and .958). Beta weights for mean sports and games ratings were .78 and .42 for sports which are games, and .77 and .44 for games which are sports. Residual plots showed that the equation systematically underestimated the ratings of very good and very bad conjunction members, compared with those in the middle of the scale. This residual variance was captured with a significant interaction term defined as the product of the sport and game ratings. New beta weights for sports and games ratings were now .67 and .40, and .70 and .42 for the two equations, and for the interaction term was .24 and .15 (R = .974 and .965). Although accounting for only a small increase in the variance explained, the interaction term removed the systematic deviation of items from the predicted values. It implies that with suitable linear rescaling, a *multiplicative* averaging function could be used to derive conjunctive membership (see Oden, 1977).

The good fit of both a simple weighted average, and a geometric average in predicting conjunctive membership ratings emphasizes the compensation between constituent membership values in determining conjunctive membership. This compensation was a major source of the overextension observed. The difference in weights attributed to sports and games is mysterious and was unexpected; the difference between the two regression coefficients was significant in the simple weighted average equations, and nearly significant in the multiplicative equations. Experiments 3 and 4 provide further evidence of this curious asymmetry.

Individual response analysis. Table 3 shows mean conjunction ratings for all subjects tabulated for each possible combination of ratings for sports and games. The process of compensation is clear, with mean conjunction ratings declining almost monotonically along each row and down each column. The top right and lower left quadrants show the overextension of the conjunctions, where with one constituent +3 and the other −1, the conjunction is still on average positive. The dominance of sport over games is also apparent in the generally higher numbers in the top right quadrant as compared with the equivalent entries in the lower left.

Commutativity. The final analysis compared the two conjunctions "sports which are games" and "games which are sports." Although logically equivalent, they may yet differ psychologically. A random split-half correlation test was made to

Table 3

Mean Ratings of the Conjunction for Each Value of the Constituent Rating Responses: Experiment 2

Sports	Games						
	+3	+2	+1	0	−1	−2	−3
+3	2.53	1.30	2.04	1.18	1.34	0.57	0.64
+2	1.54	1.56	1.12		0.79	−0.50	−0.37
+1	0.94	1.29	1.26		−0.22	−0.84	−1.60
0	2.13			−0.50			
−1	0.44	0.94	−0.05		−1.00	−2.00	−2.47
−2	−1.29	−0.70	−1.83		−1.71	−2.00	−1.89
−3	−1.97	−1.79	−2.10	−2.40	−2.00	−2.44	−2.41

Note. Only points with $n > 4$ are shown.

see whether items might be consistently better members of one conjunction or the other. The subjects were randomly divided into two groups and the differences between "sports which are games," and "games which are sports" mean ratings across items were correlated between the two groups. The correlation although low was significantly positive, $r = 0.354$, $n = 43$, $p < .01$, one-tailed, indicating that some items were consistently more typical as "sports which are games" than as "games which are sports." However, no principled account of this pattern was obvious. A more powerful test of the basis of noncommutativity was made in Experiment 4.

Experiment 3

Experiment 2 produced three interesting results. First, there was compensatory overextension of the conjunction. Second, both weighted average and multiplicative average functions provided excellent fits in predicting mean conjunctive membership. Third, there was a marked asymmetry between sports and games in their relative influence in predicting mean membership rating in their conjunction. Experiment 3 was essentially a replication study, in which the factor of the two versions of the conjunction was replaced by another experimental manipulation designed to test for a possible account of the overextension based on a range effect.

Given a list of items, half of which are sports and (a different) half games, then at Stage 1, roughly half of the list should receive a positive rating for each concept. At Stage 2, however, only about a quarter of the items may be expected to be in both categories. Therefore, if subjects try to use the full range of the scale, giving equal numbers of positive and negative responses, they

---

[7] Using multiple regression to assess different quantitative mapping functions relies on the assumption that the rating scale forms an interval scale and that judgments of item typicality (positive ratings) and item relatedness (negative ratings) can be sensibly equated. A more exact test could have used estimated scale values for the response scale (Oden, 1981) to provide a more sophisticated analysis of the mapping function. However, because the aim of the analysis was chiefly to demonstrate that conjunctive membership is a monotonically rising function of constituent membership, the exact shape of the mapping function was not pursued further.

may adopt a lower criterion when judging conjunctive membership simply in order to increase the relative number of positive ratings, and would therefore overextend the conjunction. This account is rendered more plausible by the fact that in Experiment 2, twelve mainly positive items were dropped from the list between rating the constituents at Stage 1 and the conjunction at Stage 2.

To test for this range effect, Experiment 3 included an *augmented* condition in which the list of items for rating at stage 2 was constructed by taking the Stage 1 list, removing 8 items that would not have belonged in the conjunction, and inserting 17 new filler items which were all good members of the conjunction. In this way the range effect (if there was one) would be avoided. Half of the subjects served in the augmented condition and half had the same list repeated in both stages (the standard condition). Comparison between the groups was made for a set of 36 test items which were rated at both stages by both groups. This manipulation has additional theoretical interest. If membership judgments can be affected by list context in this way, this would imply an extremely flexible category criterion for conjunctions. It would also provide support for an interesting suggestion by Jones (1982) for an extensional model of conjunctions. In Jones' model, constituent membership values are multiplied together, and items are rank ordered by the resulting values. The rank orders are then rescaled to produce the conjunctive membership scale. Presumably then, loading the list with a large proportion of new filler items, all of high typicality in the conjunction, should have the effect of depressing the mean ratings of the remaining category members.

Method

The design employed two between-subject factors: the augmented versus standard list conditions, and the order of rating the constituent concepts at Stage 1. As for Experiment 2, items were rated in Stage 1 as sports and as games, and then in Stage 2, exactly 2 weeks later, they were rated for the conjunction. All subjects rated the same conjunction, "sports which are games." Fifty-five students completed both stages of the procedure in a class setting. In addition, 12 students just completed Stage 2 (7 in the standard and 5 in the augmented condition), and so provided a control for the effects of initial testing at Stage 1.

Results and Discussion

Inconsistency and overextension. Table 1 shows the frequencies of the eight different triples of positive and negative responses across all subjects. The numbers of inconsistent [+−+] and [−++] triples replicated Experiment 2 closely. Also shown in Table 1 are the expected frequencies based on the stochastic model. In this case, mean estimated u was .527. As before, there were more overextensions than was predicted by the model. Appendix C lists those items showing most overextension. All show strong membership in one category ($>.92$) and marginal membership in the other (between .1 and .87) with the conjunction value lying in between.

Analysis of variance of the number of overextensions showed a main effect of type of overextension, $F(1, 51) = 8.49, p < .01$, with more [+−+] ($M = 3.93$) than [−++] ($M = 2.38$). The contrast effect interaction was marginally significant, $F(1, 51) = 3.94, p = .05$), and confined to an effect on the games ratings

alone. No other effects were significant. Notably, there was no significant effect of augmenting the list on the amount of overextension observed in the constant set of 36 items. The average numbers of items rated positively for both constituents at Stage 1 were 12.8 for the augmented group, and 12.4 for the standard group. At Stage 2, the augmented experimental group gave 17.6 positive ratings (and the control group 17.2). The standard experimental group gave 19.1 (and the control group 16.7). None of the differences were significant.

Mean item ratings. Regression analysis was applied to mean item ratings (combining both subject groups). The weighted average function had beta weights of .82 for sports and .44 for games ratings ($R = 0.952$). Again the weighted average underestimated items belonging to both or to neither category, compared with those belonging to just one. This systematic deviation was captured by a significant interaction term (beta weights: sports .72, games .51, and the interaction .27; $R = .986$). The weight for sports was again significantly greater than that for games, this time both with and without the interaction term in the equation. This imbalance was confirmed in the response matrix equivalent to Table 3 (not shown), in which 3 for sports guaranteed a positive mean rating for the conjunction, whereas 3 for games did not.

Conclusions. The results of Experiments 2 and 3 agreed closely. In both it was found that either a weighted average, or a multiplicative function predicted mean ratings in the conjunction extremely well. The conjunction was overextended in a way that was not solely due to contrast, range effects, or subject variability but seemed to be due to the application of compensation to the determination of conjunctive membership. The absence of any effect of list augmentation on membership in the conjunction lends no support to Jones's (1982) rescaling model of conjunctions.

The final experiment extended the results of Experiments 2 and 3 to a wider group of concepts. Of particular interest were (a) the generality of the good fit of the regression equations (b) the unequal weights of sports and games in those equations, and (c) the question of noncommutativity.

Experiment 4

In Experiments 2 and 3 a highly systematic mapping relation was found to exist between constituent and conjunctive membership, in spite of Osherson and Smith's (1982) theoretical arguments against the generality of such a function. Experiment 4 examined whether this mapping function, might be in some way specific to the concepts of *sports* and *games,* by using six new conjunctive concepts. Experiment 2 left unanswered the question of how the two phrases "A which are B" and "B which are A" may differ in their extensions. Although consistent differences between the two conjunctions were found, no obvious explanation was apparent. A new set of categories allows a more sensitive test of this problem. Finally, the curious dominance of sports over games in predicting conjunctive ratings for Experiments 2 and 3 needed to be explored in other concepts. Was the imbalance peculiar to these two concepts, or is it the norm for one concept to predominate in forming a conjunction? If it is usual, what distinguishes the dominant member of each pair?

JAMES A. HAMPTON

Table 4
Distribution of Response Triples for Conjunctions in Experiment 4, Showing Percentage of Inconsistent Responses of Each Type

Conjunction	Response triple [A, B, conjunction]							
	$+++$	$++-$	$+-+$	$+--$	$-++$	$-+-$	$--+$	$---$
A. Furniture B. Household Appliance	132	22	81	173	111	65	2	51
% inconsistent responses		13 (18)		32 (8)		63 (9)	4 (6)	
A. Food B. Plants	291	15	60	106	25	75	4	27
% inconsistent responses		5 (5)		36 (4)		25 (13)	13 (2)	
A. Weapons B. Tools	227	24	79	59	74	130	5	38
% inconsistent responses		10 (9)		57 (15)		36 (6)	12 (1)	
A. Buildings B. Dwellings	321	27	30	84	46	51	6	71
% inconsistent responses		8 (7)		26 (12)		47 (23)	8 (2)	
A. Machines B. Vehicles	291	25	35	84	34	91	3	68
% inconsistent responses		8 (6)		29 (7)		27 (10)	4 (2)	
A. Birds B. Pets	254	12	57	99	1	148	0	53
% inconsistent responses		5 (4)		37 (8)		1 (0)	0 (0)	

Note. Values in parentheses are those predicted by the stochastic model.

Method

Design. The design was identical to that of Experiment 2, with the two independent group factors, order of rating the two constituents at Stage 1 and order of concepts within the conjunction.

Materials. Six pairs of overlapping categories were chosen. They were: *machines–vehicles; furniture–household appliances; pets–birds; buildings–dwellings; food–plants; weapons–tools.* After piloting, the concept *plant* was changed to "*plant* (or part of a *plant*)" to reduce an ambiguity in the *plant–food* conjunction. For each conjunction 16 items were selected, including items in both sets, in either one or the other, and in neither. Borderline examples were included in order to spread items along the scale. There was a slight preponderance of items in both sets to discourage range effects (although Experiment 3 had found no evidence for such effects) and to provide a range of conjunctive typicality.

Procedure. The procedure was the same as before. In Stage 1, subjects rated 12 concept lists, corresponding to the two constituent concepts in each conjunction. Lists were printed 2 to a page in a booklet, each concept being paired with one from a different conjunction. Order of items within lists was random for each concept. Page order was balanced across subjects, equal numbers doing each page in each ordinal position. Thus, assignment of subjects to the order-of-rating factor was random and different for each conjunction. In Stage 2, one week later, subjects rated six lists, one for each conjunction. For each order of rating at Stage 1, half of the subjects now did "A that are B," and the other half did "B that are A." Order of the six lists was again balanced. Subjects were run in small groups. One subject failed to return for Stage 2 and was replaced.

Subjects. A total of 40 students (10 in each condition) taking an introductory psychology course at Stanford University participated for course credit. All were fluent speakers of English.

Results and Discussion

Inconsistency and overextension. Table 4 shows the frequencies of each type of response triple, together with the observed and expected percentages of inconsistent triples. The same stochastic model was applied. The mean recall parameter u was .74, higher than before, perhaps because of the shorter delay between the two tests. The number of overextended conjunctive responses was again high, with the notable exception

of *birds.* No items were judged to be in the conjunction which were not previously judged to be *birds,* (although they might not be *pets*).

The contrast between constituent ratings due to repeated testing of the same list, found in Experiments 2 and 3, was reduced in this experiment where each pair of lists was embedded in a total set of 12 lists. Of the 12 concepts, 9 had fewer positive ratings when rated second, but none of the differences were significant. The ANOVA for each concept pair analysed the number of $[+-+]$ and $[-++]$ response triples, with order of testing at Stage 1, and order of concepts within the conjunction as between groups factors. In the six ANOVAs there were two significant effects of type of inconsistency (because of low levels of overextension for *birds* and *food;* see Table 4). One Type × Conjunction Order interaction was significant; good furniture was rated better as "*household appliances* that are also *furniture,*" whereas good household appliances did better in the converse conjunction. One three-way interaction was significant but had no clear interpretation and no other effects were significant. No contrast effect interactions were significant. Appendix D lists the examples with most overextended response triples in each conjunction. As before, they tended to be good members of one category, and marginal members of the other.

Mean item ratings. Mean ratings for each item in each concept and conjunction were again calculated. In only 7 of the 96 items did the rating for the conjunction exceed both constituents, and in each case the difference was only marginal. The guppy effect seems hard to find. For instance, a close parallel to guppies might be expected in the case of the *pet bird* category. The best *pet birds* were *canary* (2.949), *parakeet* (2.9), and *parrot* (2.725). However, these were also very typical as *birds* (*M*s = of 2.95, 3.0, and 2.95 respectively, when rated in initial position).

Although *birds* were immune from overextension, this could not be attributed to a dissociation of typicality judgments from class membership decisions. It was apparently caused by a lack of variation in the rated typicality of *birds* in this study. All the birds had mean ratings greater than 2.25, and all nonbirds had

mean ratings less than −2.7. It happens that in the knowledge domain of creatures, birds form a clear-cut category (a point noted previously, Hampton, 1979), so there are simply no suitable candidates for overextension. (It seems unlikely that this clarity extends to all lay biological categories. *Fish, insects, animals, flowers,* and *trees* all have borderline cases; see Hampton & Gardiner, 1983.)

As an incidental result, a further test of Jones's (1982) model could be made. His model requires that the best member of a conjunction should always be rated at the top of the typicality scale. However, in the conjunction "*furniture* that is also a *household appliance,*" the best examples in the conjunction—television, cooking stove and hi fi—all received only moderate mean typicality ratings (less than 2.0). A subsequent task in which subjects were asked to generate members of the category failed to come up with any exemplars more popular than these. It seems that rescaling within conjunctions need not necessarily occur.

Regression equations were used to predict mean conjunction ratings from mean constituent ratings. A total of 12 equations were computed, one for each version of each of the six conjunctions. The weighted average function provided *R*s between .905 and .971. As in Experiments 2 and 3, a small positive interaction term entered the equation significantly in 10 of the equations, which raised the *R*s to between .931 and .995, with a mean level of .97. Table 5 shows the regression statistics for each equation.[8] Several equations again showed bias towards one constituent. *Weapons, household appliances, dwellings,* and *birds* were dominant over *tools, furniture, buildings,* and *pets,* respectively. (Two-tailed *t* tests for the differences in regression coefficients were significant for both *pet bird* equations, and for "*buildings* that are also *dwellings*." With only 16 items per category, the power of these tests was obviously much lower than for earlier experiments.) For *food–plants* and *machines–vehicles,* the weights were very similar when interaction terms were included, although without them, *plants* and *vehicles* had higher weights. The same dominance pattern appeared in both versions of each conjunction. However, the nondominant category was not always the most overextended. The effect appears to be independent of the distribution of examples used in the lists. (Variance did not differ significantly between dominant and nondominant concepts.) The sources of this dominance effect will be discussed in the final section.

Noncommutativity. Table 5 also shows the effects of changing the order of concepts within the conjunction. For 10 of the 12 concepts, beta weights were higher when the concept was in the relative clause than when it was the head noun. Thus, over and above the dominance effect, a concept had more influence on a conjunction when it was in the qualifier clause. Mean ratings confirmed this pattern. Of items with different values for the two forms of the conjunction, 64% were rated higher when the concept for which they had a higher rating was in the relative clause, ($z = 2.57, p < .01$ on a binomial test). It was also notable that with the *dominant* concept as head noun, the weight of the interaction term was increased.

Individual response data. Because there was no significant contrast effect in this experiment, a more elaborate analysis was made of the 7 × 7 matrix of mean conjunction ratings for each possible combination of constituent ratings. Table 6 shows the

matrix, collapsed across all 12 concepts, with rows corresponding to head-noun rating, and columns for relative clause-noun rating. The data were used to construct a graph showing contours of equal rated membership in the conjunction as a function of membership in the constituents.[9] Figure 1 shows lines joining points of equal rated membership in the conjunction at 0.5 intervals from −2.5 to +2.5. The vertical axis shows membership in the head-noun concept (A), and the horizontal axis membership in the relative clause-noun concept (B). The additional weight given to the qualifier noun is shown by the tendency of the lines to slope down with a slope steeper than −1.

Although the scales are not ratio scales, the data can be compared with three functions. First, a weighted average would give straight lines from the edges in towards the positive diagonal. (Dominance effects would not show up because all 12 categories contribute to both axes.) The lines are indeed quite straight and parallel above the major diagonal (where membership is greater in the head noun than in the qualifier noun), but in the lower right half of the graph they curve quite sharply. Second, a multiplication function (with rescaling) would yield a family of curves, concave towards the upper right corner. The lines fit this description better for values below the diagonal. Finally, a third function, Zadeh's minimum rule for conjunction would give rectangular L-shaped lines. Clearly this function can be ruled out as an account of the data. One possibility suggested by the graph is therefore that subjects' decision rule may depend on whether an item is a better member of the head-noun or the qualifier-noun class. In summary, the weighted average function appears to fit the top half of the diagram better, whereas a multiplicative function may fit the lower half; however, transformation of the scales could alter this pattern.

The main point for the distinction between unitary and binary views is that both functions involve compensation and overextension. The boundary of the conjunction in Figure 1 (the line labeled *0.0*) passes close to the origin of the two scales, in the centre of the diagram. This suggests that subjects may set the criterion of what can belong in a conjunction (as they choose to define it) by relaxing the criterion to the point where to be more inclusive would allow items to enter which were in *neither of the constituent sets.* This they do not appear willing or likely to do, and seems to be a firm constraint on what may belong in a fuzzy conjunction.

---

[8] The highly predictable relation between constituent and conjunctive scale values suggests that in practice there may well be relatively simple functions that map the one on to the other. However, theoretical arguments (Osherson & Smith, 1982) against any general function cannot be ignored, and they advise caution in seeking to refine the quantitative modeling of the data. The paradoxical conflict between data and theory may perhaps be resolved by the development of an intensional model for concept conjunctions (see discussion later).

[9] In the extreme upper left and lower right corners, the lines reverse their slopes. The likely explanation for this is a familarity confound. Items rated [+3 −3] are more likely to be familiar than those given a more cautious [+2 −3], and so receive a more extreme negative rating for the conjunction, in spite of having apparently better positive membership in one concept.

Table 5

Regression Equations Predicting Mean Ratings of Items as "A That Are Also B" From Mean Ratings as A and as B: Experiment 4

		A that are also B				
		Regression weight				
A	B	A	B	A&B	Intercept	R
Furniture	Household Appliances	.481	.850	—	−.567	.957
Beta weight		.493	1.214	—		
Household Appliances	Furniture	.634	.559	.122	−.526	.931
Beta weight		1.079	.683	.246		
Food	Plants	.526	.479	.183	−.522	.973
Beta weight		.546	.467	.441		
Plants	Food	.487	.532	.191	−.536	.973
Beta weight		.463	.538	.449		
Weapons	Tools	.619	.501	.124	−.643	.960
Beta weight		.854	.534	.225		
Tools	Weapons	.549	.805	—	−.468	.963
Beta weight		.553	1.051	—		
Buildings	Dwellings	.337	.712	.073	−.391	.982
Beta weight		.351	.740	.162		
Dwellings	Buildings	.577	.435	.130	−.712	.965
Beta weight		.590	.445	.281		
Birds	Pets	.602	.445	.140	−1.053	.995
Beta weight		.766	.347	.349		
Pets	Birds	.411	.648	.112	−.938	.990
Beta weight		.320	.823	.279		
Machines	Vehicles	.468	.598	.118	−.728	.984
Beta weight		.492	.638	.304		
Vehicles	Machines	.540	.553	.128	−.633	.963
Beta weight		.553	.559	.317		

General Discussion

Synopsis of Results

In this section, the results will be summarized with a view to likely boundary conditions and other possible limitations on their generality. One central aim of this study was to test the idea that a unitary basis of typicality and class membership was possible for some natural concepts. The data certainly supported this idea. Osherson and Smith (1982) correctly pointed out that unitary theories of concept conjunctions predict logically inconsistent decisions about set membership in conjunctions. If membership in a conjunction rises monotonically with degree of membership in each concept, then compensation and hence inconsistency are inevitable. However, the results showed that the prediction of logically inconsistent responses to conjunctions may be a virtue rather than a failing. In at least one kind of categorization, subjects did allow compensation and did make inconsistent classifications.

Various artifactual sources of overextension were examined. A contrast effect was found, in which ratings for a category were depressed if the same items had just been classed in another category. The variability of categorization was also taken into account, and a range effect on the rating scale was experimentally tested. However these three artifacts were not sufficient to explain the high degree of overextension observed.

Regression analyses showed that a substantial amount of overextension was owing to the subjects' responses obeying a compensatory averaging function (either linear or geometric) to

map conjunctive ratings from constituent ratings. Equivalent effects of overextension were found from analysis of probability of class membership, and analysis of mean rated typicality. This equivalence supports the unitary hypothesis. If people tried to be consistent in their class inclusion decisions, then overextension should occur mostly in the analysis of mean typicalities.

It is feasible that something in the nature of the rating task led subjects to an averaging rather than a logical rule. Combining both a typicality judgment and a class membership decision on a single scale may have had this effect, as may the fact that they had to make around 90 different judgments in the same

Table 6

Mean Ratings of the Conjunction for Each Value of the Constituent Rating Responses: Experiment 4

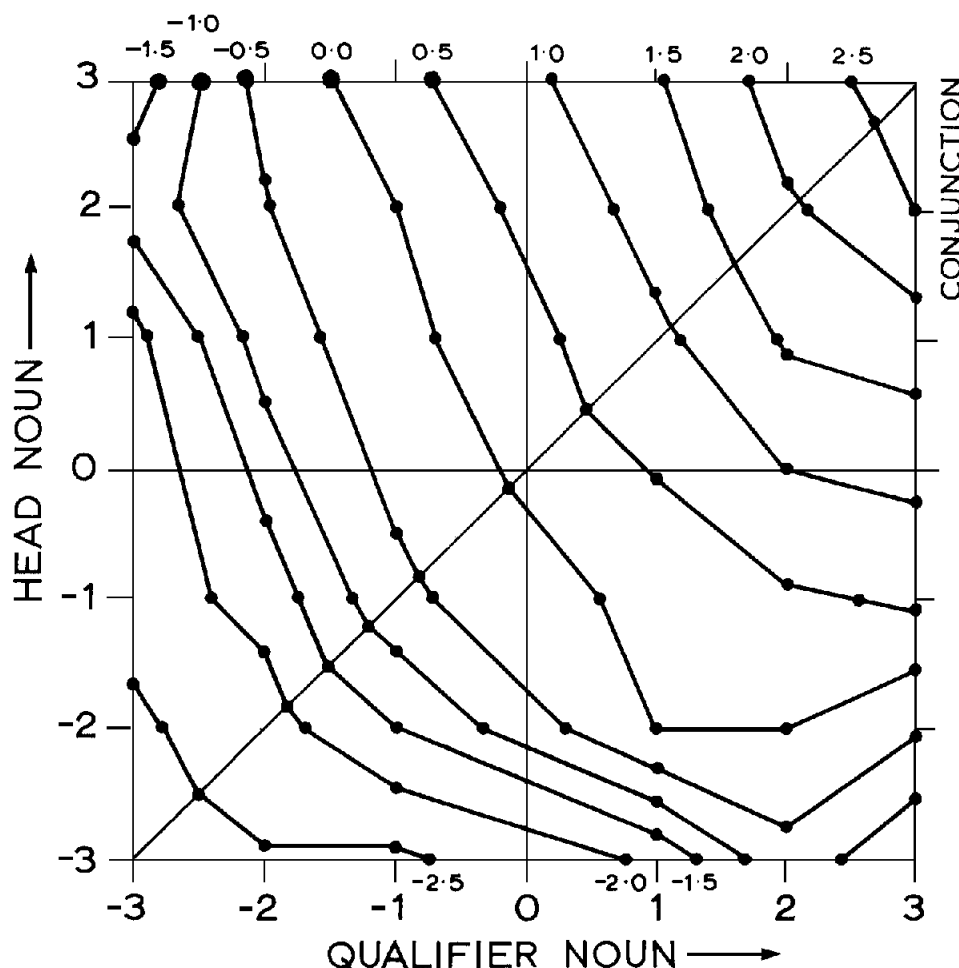|	Qualifier noun						
	+3	+2	+1	0	−1	−2	−3
Head noun							
+3	2.75	2.24	1.47	0.00	0.38	−0.37	−1.76
+2	2.51	1.94	1.21	—	0.00	−0.51	−1.26
+1	1.71	1.58	0.89	0.58	−0.13	−0.75	−2.17
0	1.50	0.12	−0.67	1.00	—	—	—
−1	0.60	0.43	0.18	0.25	−0.60	−1.81	−2.32
−2	−0.43	0.02	0.03	—	−1.50	−2.24	−2.59
−3	−1.45	−0.67	−1.89	−1.33	−2.59	−2.54	−2.76

Note. Only points with $n < 4$ are shown.

Figure 1. Lines of equal membership in the conjunction drawn for different values in the head-noun category (vertical axis), and the qualifier-noun category (horizontal axis), summed over all responses in Experiment 4.

session. Only further research can tell under what circumstances subjects can be led to treat conjunctions extensionally. It would be most unlikely that they never do so. For the moment, the present data suggest that there is at least one set of conditions under which subjects are apparently happy to ignore Boolean logic. For example, on the final question sheet in Experiment 1 subjects were happy to agree that "school furniture is a type of furniture," while at the same time claiming that "some school furniture is not furniture."

Two other new and interesting findings emerged from the experiments. First, it was shown that although extensionally equivalent, the two phrases "*tools* which are also *weapons,*" and "*weapons* which are also *tools,*" were not treated the same. Experiment 2 failed to show a systematic basis for this asymmetry in the cases of "*sports* which are *games,*" but Experiment 4 showed clearly that greater importance is given to the concept in the relative clause–qualifier position. All six conjunctions showed the same effect, although to different extents.

The second interesting finding was the dominance effect between the two constituents of a conjunction. In several cases the

imbalance between the two concepts was extreme. Because this result was unexpected, its source may be hard to identify. It is perhaps possible that the effect is an artifact of rescaling membership values on entering them into a multiplicative function, or else some effect of the distribution of items across the two membership scales. Against an artifactual account one can offer the evidence in Table 3, which shows how the compensatory values of *sport* and *game* were very different in Experiment 2. Similar tables could have been presented for Experiments 3 and 4 showing the same effect, so at the least it can be argued that the effect is not specific to the use of a regression analysis. A second argument for the validity of the effect is that in two replications using six of the present seven conjunctions with a between-subjects design, and new distributions of items, the dominance patterns between concepts were for the most part confirmed. Thus, if it is an artifact, it is a very reliable one and in need of an explanation.

The theoretical importance of the dominance effect, if it should prove valid, is that it emphasises the interactive nature of conjunction. It is unclear how a simple extensional context-

free rule could explain such an imbalance between concepts. At some point, comparative information about the two categories—not just the degree of membership of an item in each one separately—must be introduced to the rule. Such information could be extensional (e.g., the distribution of category exemplars) or intensional (the average dissimilarity of category exemplars, or the number of salient attributes defining the category prototype). In either case, concept dominance would provide an important phenomenon against which to test different models of conceptual combination.

Composite Prototypes and Intensional Combination

This section describes a model of concept conjunction based on the notion of a *composite prototype* (Hampton, 1983), which offers an account of the results obtained. Following discussion of the model and the theoretical issues involved, the final section returns to the explanation of the empirical findings and considers alternative models.

The general approach to concept combination taken here has had several advocates (Cohen & Murphy, 1984; Hampton, 1983; Smith & Osherson, 1984; Thagard, 1983). The central tenet is that successful prediction of conjunctive class membership and typicality does *not* depend on finding an operator to map constituent concept extensions on to their conjunction's extension (as fuzzy logics aim to do: Turksen, 1984; Zadeh, 1965), but must instead provide rules for constructing an *intensional prototype* for the conjunction itself. The conjunction's extension can then be defined using the same mapping function from intension to extension as is used in prototype theory for simpler noncombinatorial concepts.

It may seem strange to adopt this approach in the light of the apparently highly systematic extensional mapping function seen in the data. For example, it appeared from Figure 1 that subjects might be using an arithmetic average function for items that are better members of the head-noun class, and a geometric average for those that are better members of the qualifier class. There are two main reasons for eschewing an extensional mapping solution, one theoretical and one empirical.

First, Osherson and Smith (1981, 1982) argued that a single function cannot work for all conjunctions. For example, Smith and Osherson (1984) showed empirically that when two concepts are positively associated (as in red apple), a very different mapping function applies from the case where they are negatively associated (as in striped apple). Furthermore, taking the analysis of their pet fish example further, it is possible to show that the guppy example breaks a general property of extensive mapping rules known as *independence* (Krantz, Luce, Suppes, & Tversky, 1971). This property states that for any item x and for any four concepts A, B, C and D, and their membership functions C,

$$C_{AB}(x) > C_{AC}(x) \text{ iff } C_{DB}(x) > C_{DC}(x).$$

Suppose that x is the guppy, and the four concepts are A = *domestic*, B = *fish*, C = *pet*, and D = *aquatic*. Then we have (all for the case of *guppy*):

$$C_{\text{domestic fish}} > C_{\text{domestic pet}}$$

iff

$$C_{\text{aquatic fish}} > C_{\text{aquatic pet}}.$$

Because intuitively a *guppy* is a highly typical domestic fish and aquatic pet, and is not particularly typical of domestic pets or aquatic fish, it seems very likely that this rule would be broken by subjects' typicality judgments. (A related demonstration involving square and round circles is offered by Osherson and Smith, 1982).

The empirical reason for preferring an intensional model is the appearance of the dominance effect in Experiments 2, 3, and 4. Most simple extensive mapping functions could not motivate the incorporation of different weights for the two constituent members of a conjunction (assuming that they prove reliable). It is possible for more sophisticated functions to be developed, but they would have to be able to predict which concept dominates. One obvious extensional candidate for predicting dominance—set size—does not work. Rough estimates of relative set sizes can be obtained by asking which is the greater, "A that are not B," or "B that are not A"? It seems that there are more birds that are not pets than pets that are not birds (both in types and in numbers of individuals). Similarly, there are more buildings that are not dwellings than there are dwellings that are not buildings (at least around Stanford). However although birds and buildings are the two larger categories on this criterion, it is birds and dwellings which dominated their conjunctions. Given these reasons for not pursuing an extensional account, the remaining part of this section develops an intensional model of conjunction.

The development of an intensional model depends on the notion that conjunctions "inherit" attributes from their constituents in an *interactive* way. It then becomes possible for an item's similarity to the conjunction to be greater than its similarity to either constituent. To illustrate this with the guppy example the conjunction *pet-fish* shares some of the attributes of pets (*domestic habitat, cared for*), and some of fish (*cold-blooded, no limbs*). At the same time there are other attributes of each constituent which are *not* inherited (e.g., *warm and cuddly* for pets, *lives in the sea* for fish). The result is that *guppy,* which happens to match all the attributes of *pet-fish,* will not match all the attributes of the constituents *pet* and *fish,* and so will be more typical in the conjunction than in either constituent. The critical issue is how to predict which attributes of concepts will be inherited by conjunctions and which will not. Why are pet fish not warm and cuddly creatures living in the wild? (Some empirical work toward answering this question can be found in Hampton, 1987). The following five subsections consider different possible rules for attribute inheritance by conjunctions. It would be premature to attempt a full model of composite prototype formation. The following proposals are speculative and not mutually exclusive.

Specified versus default values. Thagard's (1983) approach to concept conjunctions used frame theory (Minsky, 1975). Semantic features are represented as slots (such as *color*) which can take various slot fillers or values (*red, green*). Thagard proposed that where a slot filler is *specified* it will be inherited in

preference to others that are merely *defaults*. For example, *flightless birds* must have the value [does not fly] for *locomotion* because the specified value for *flightless* will overwrite the default value for *birds*. Smith and Osherson (1984) developed a similar model for adjective–noun pairs (see also Smith et al. 1985). These proposals seem noncontroversial for making typicality judgments about objects such as *red apples*, which is the intended scope of their model. However, no obvious way suggests itself for extending the principle to cover noun–noun conjunctions such as *pet fish*. If "specified" means "explicitly stated," then none of the attributes of *pets* and *fish* are specified, except for petness and fishness. Additional rules are therefore needed.

Diagnosticity. A second rule (Thagard, 1983) with a wider range of application is that inheritance will depend on selecting the slot filler that has highest diagnosticity for its parent concept. Presumably, specified values have maximal diagnosticity, so this rule incorporates the previous one. The value of this rule has yet to be fully tested empirically, partly because of the lack of an uncontroversial definition of diagnosticity. However, evidence from one study (Hampton, 1987) suggests that the perceived importance of an attribute for a conjunction is a weighted average of its importance for each constituent concept, subject to two constraints. First, attributes that are considered necessary for either constituent are also considered necessary for the conjunction. Second, attributes that are considered impossible for either constituent are also considered impossible for the conjunction. These constraints can in fact be formalized[10] in a function derived from the Dempster-Shafer rule for combining uncertainty (Ginsberg, 1984).

Context dependence. Two other rules proposed by Thagard (1983) may be subsumed under the notion of context dependence. The first is that attributes that are less relevant to the conjunction in the context in which the concept is used may be deleted from the representation. The second rule is that two default values may both be inherited, so that the context (for example the object being categorized) can determine which value is used to represent the conjunction. By this rule, the overlap of attributes between an exemplar and a category would be maximized within the constraints of variation of the category prototype. There are few data on either of these rules. Dropping irrelevant attributes provides another way in which an item may have greater overlap with a conjunction than with its constituents. Choosing attribute values to maximize overlap between an example to be categorized and the category representation would match well with other results in similarity research (Tversky, 1977).

Mental models and theories. Several authors have recently argued for the need to consider concepts as embedded in, and embodying our theories about the world, (Cohen & Murphy, 1984; Lakoff, 1987; Murphy & Medin, 1985). They argued that it is dangerous to take a too simplistically compositional approach to representing concepts, and the concept-as-theory approach matches other work on semantic representation using "mental models" of the world (Johnson-Laird, 1982). When forming conjunctions of concepts, this view emphasizes the need for the conjunction to be *coherent*. Thus, one principle for deciding which of two slot values should be inherited is to select the value that makes the most coherent connections with the

other attributes of the concept. Indeed, to increase coherence, some attributes of conjunctions may be *inferred* and not inherited at all. For instance (Murphy, 1987), *green* artifacts tend to be *painted*, whereas *green* fruits tend to be *unripe*. These are attributes inferred from theories about the source and significance of the color of objects. The idea of concept-as-theory almost certainly can provide insights into the problem of conceptual combination, although as yet it is insufficiently precise. This may reflect difficulties in representing the structure of theories. Hampton (1987) presented some preliminary data that support the view. In its most radical expression the approach would predict that as individual concepts each consist of richly detailed and different kinds of semantic information, the formation of composite concepts will never be straightforward, and may not conform to general principles (see also Armstrong et al., 1983). To the extent that the present experiments show a fair degree of uniformity among different conjunctions, the influence of noncompositional effects may be limited.

Extensional feedback. Another way in which attributes can come to be associated with a composite prototype is through our *factual* (as opposed to theoretical) knowledge of things in the world. Once a composite concept has been formed, those objects that most closely match that concept can be identified. Further regularities and diagnostic attributes for this class of objects, may then be recognized and included in the general intensional representation. In this way additional information can be acquired, through *extensional feedback*, to flesh out the conjunction's intension. (To stretch the genealogical metaphor one might call such attributes *acquired*.) Some examples will illustrate this point. Pet birds are kept in cages, and sometimes talk, although pets and birds have neither attribute (see Hampton, 1987). Aquatic mammals have blowholes for breathing, although this is not generally true of aquatic creatures, or of mammals. These things have been learned about the world, and come to be attached to the particular conjunctive concepts. Another case is the way a noun concept modifies the value of a qualifier. For example, different hues are typically associated with *red* when it is applied to wine, hair, apples or tulips. What we know about objects influences conjunctive intensions, and hence indirectly affects judgments of conjunction typicality.

Accounting for the Empirical Findings

Following the theoretical discussion of attribute inheritance rules, the final section discusses the experimental results in relation to the proposed intensional approach and considers alternative explanations.

Overextension of conjunctions. According to classical set theory, the intension of a conjunction is formed by the *union* of the necessary criteria of the two constituents. Thus, in terms of the present model, inheritance of attributes is complete. Competition between rival slot fillers would render a conjunction empty, that is, a logical impossibility. If the strict "common element" definition of concepts is relaxed, then intensional composition of a conjunction must necessarily involve the creation of a new prototype composed at least in part of (selected) attri-

---

[10] The formalism was suggested by Mark Gluck.

butes of the two constituents. As a result, similarity of any item to the conjunction will increase as a monotonic function of its similarity to either constituent, just as was seen in the typicality ratings in Table 3. In effect, this means that the conjunction must be either over- or underextended, because of compensation between the two membership values. The results suggested that faced with the choice of over or under extending a conjunction, most subjects maximized membership in the conjunction by placing the criterion for membership as wide as possible (see Figure 1).

The two quantitative functions suggested by the data can be related to the intensional model as follows. A weighted average function implies that the two sets of attributes are simply overlaid in a noninteractive fashion, summing the relative importance of each attribute for the two concepts. A geometric average would result from interactive and noncompositional effects of the kind described in the previous section because their effect is to increase the fit to the conjunction of items that are members of both constituent sets, relative to those that are in only one. The lines in Figure 1 suggest that interactive effects are greater when an item is a better member of the qualifier concept than of the head noun concept.

Neither averaging function would of course account for the *guppy effect* because averages always lie between the two averaged data points. The extent to which membership in conjunctions can rise above that in *both* constituents is unclear. Osherson and Smith's (1981, 1982) examples carry intuitive appeal, but the authors provide no direct empirical evidence (it is difficult to interpret data from adjective–noun conjunctions; see Hampton, 1983). The data of the present experiments found no evidence for the effect (hence the good fit of averaging functions). Even where concept pairs were negatively diagnostic as, for instance, *furniture* and *household appliances,* the most typical members of the conjunction did not rise significantly above the higher constituent typicality level. Further research is therefore needed to establish the conditions under which the guppy effect may appear. The lack of effect suggests that the mutual interaction of attributes exemplified by *pet fish* may be relatively uncommon in conjunction formation.

The composite prototype model is not the only possible account of overextension, and alternative accounts exist that preserve the binary hypothesis in some form. For example, subjects may use a classical conjunction rule but lower the membership threshold for one constituent, provided that an item is a good member of the other. (It would not be enough simply to lower both the thresholds because items that are just outside both constituents do not get included in the conjunction; see Figure 1.) The explanation of why they should do this is unclear, but it is certainly a serious alternative, making essentially the same predictions about overextension. Note, however, that the notion of a flexible category boundary which can be extended in a smooth fashion is inconsistent with the spirit of the binary position. If core definitions exist and are used in this task, they are unlikely to allow such elastic category boundaries. The account would also have to find an explanation for the concept dominance effect.

A second alternative account of overextension that would rescue the binary hypothesis is that although subjects use a classical conjunction rule, the nature of the representation of the two

concepts is altered by being in the *context* of a conjunction, so that the decision is based on altered semantic representations. This suggestion may be very similar to the composite prototype model, involving as it does an alteration of the attributes against which an example is judged. It would require similar mechanisms for contextual modification and an explanation of why contextual modification does not allow items that are just outside both constituent sets to be included in the conjunction.

Noncommutativity of conjunctions. The conjunctive phrases used in Experiments 2, 3, and 4 were noncommutative in that greater weight was attached to the concept in the relative noun position. No published accounts of conceptual combination predict this effect; in fact, even the composite prototype approach has little to add. A post hoc functional explanation of this result can be offered. In a phrase such as "*tools* that are also *weapons,*" the head noun, *tools,* is generally taken as the given domain of reference, whereas the relative clause provides new information to modify that domain, hence the traditional linguistic view (shown here to be false) that the qualifier picks out a subset of the head noun class. The qualifier is therefore highlighted, and the head noun is taken as background. A similar account is in terms of the natural *contrast set* of a conjunction. Rosch and Mervis (1975) showed that typicality in a category is also related to lack of similarity to the category's contrast set (for instance an item is a better *fruit* if it is less like a *vegetable*). The contrast set of "*weapons* that are also *tools*" involves negating the *qualifier* clause, as in "weapons that are not tools." Thus, the dimension of "toolness" assumes greater importance because it distinguishes items in the conjunction from those in the contrast set. Post hoc, the composite prototype model could propose that attributes from the relative clause concept are given greater importance in the conjunction and have a greater chance of being inherited by the composite prototype. Evidence on this question from the study by Hampton (1987) showed a trend in this direction, but the effect was not statistically significant.

Dominance between constituents. The final phenomenon to be explained is the imbalance between the two constituents in their influence on membership in the conjunction. A number of factors may underlie this dominance effect (assuming that it is nonartifactual). For example, although it was argued previously that set *size* itself does not predict dominance, it is perhaps the case that dominant concepts contain a more homogeneous set of exemplars (birds as compared with pets, weapons as compared with tools). Interestingly, the composite prototype model needs only one simple assumption to account for concept dominance. The assumption is that the intension for the conjunction is composed of a mixture of attributes from both constituents, but with *more* attributes from the dominant concept. Similarity of an item to the conjunction would then be more predictable from similarity to the dominant than to the nondominant constituent. In terms of inheritance rules, there would be at least three ways in which this could arise. First, the dominant concept may have a larger number of salient attributes for the conjunction to inherit; second, where there is conflict between attribute values (or slot fillers), those of the dominant concept may have higher diagnosicity and so overwrite those of the nondominant concept; third, the coherence of the conjunction may be better served by inheriting attributes from

the dominant concept than those of the other. In a study of attribute inheritance (Hampton, 1987), it was found in each of six conjunctions that the dominant concepts (as identified in the present study) had a higher number of salient attributes. There was little evidence, however, for the other suggested mechanisms.

The intensional theory of concept conjunction can therefore account quite readily for concept dominance, initially simply in terms of the number of salient attributes possessed by each concept. There is independent evidence that this is a potent dimension in semantic memory, being associated with item typicality (Ashcraft, 1978), item familiarity (Malt & Smith, 1982), and with determining the basic level of categorization (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). It may reflect *distinctiveness*, in the sense of the amount of specific information known about a class.

Conclusion

This research has established a number of interesting facts about how people form conjunctions of concepts under particular conditions. Conjunctions tend to be overextended to include items that are good members of one class, but are marginal to the other. They are rarely underextended. The phrase "A that is also B" is not treated as commutative. Instead, more emphasis is placed on the concept in the relative clause. The weight given to the two constituents in the conjunction may often be quite unequal. It has been argued that all of these results can be accounted for in terms of a process of combining the intensions of the constituent concepts into a composite intension for the conjunction. Membership in the conjunction is then determined with respect to this composite prototype representation by placing a criterion on the similarity of items to be categorized, such that a hypothetical item on the borderline of both constituent concepts would be on the borderline of the conjunction. Overextension is caused by allowing membership in one set to compensate for nonmembership in the other. Finally, the unequal weight of the two constituent concepts can be explained by reference to the amount of detailed semantic information in their respective intensions.

References

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

Ashcraft, M. H. (1978). Property norms for typical and atypical items from 17 categories: A description and discussion. *Memory & Cognition, 6*, 227–232.

Barsalou, L. (1984, October). *The instability of graded structure: Implications for the nature of concepts.* Paper presented at a conference on "The Ecological and Intellectual Bases of Categorization," Emory University.

Bellezza, F. C. (1984). Reliability of retrieval from semantic memory: Common categories: *Bulletin of the Psychonomic Society, 22*, 324–326.

Bourne, L. (1982). Typicality effects in logically defined categories. *Memory & Cognition, 10*, 3–9.

Clark, E. V. (1983). Meanings and concepts. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology: Vol. 3, Cognitive development.* New York: Wiley.

Cohen, B., & Murphy, G. L. (1984). Models of concepts. *Cognitive Science, 8*, 27–58.

Ginsberg, M. L. (1984). *Non-monotonic reasoning using Dempster's rule* (Heuristic Programming Project Report No. 84-30). Stanford, CA: Stanford University.

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

Hampton, J. A. (1981). An investigation of the nature of abstract concepts. *Memory & Cognition, 9*, 149–156.

Hampton, J. A. (1982). A demonstration of intransitivity in natural concepts. *Cognition, 12*, 151–164.

Hampton, J. A. (1983). *A composite prototype model of conceptual conjunction.* Unpublished manuscript, The City University, London.

Hampton, J. A. (1984, January). *Evidence for a two process model of categorization decisions.* Paper presented at the Experimental Psychology Society Conference, London.

Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition, 15*, 55–71.

Hampton, J. A., & Gardiner, M. M. (1983). Measures of internal category structure: A correlational analysis of normative data. *British Journal of Psychology, 74*, 491–516.

Johnson-Laird, P. N. (1982). *Mental models.* Cambridge: Cambridge University Press.

Jones, G. V. (1982). Stacks not fuzzy sets: An ordinal basis for prototype theory of concepts. *Cognition, 12*, 281–290.

Kahneman, D., & Tversky, A. (1972). Subjective probability: a judgment of representativeness. *Cognitive Psychology, 3*, 430–454.

Keil, F. C. (1986). The acquisition of natural kinds and artifact terms. In W. Demopolous & A. Marras (Ed.), *Language learning and concept acquisition* (pp. 133–153). Norwood, NJ: Ablex.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. 1). New York: Academic Press.

Lakoff, G. (1987). *Women, fire and dangerous things.* Chicago: University of Chicago Press.

Levi, J. (1978). *The syntax and semantics of complex nominals.* New York: Academic Press.

Malt, B. C., & Smith, E. E. (1982). The role of familiarity in determining typicality. *Memory & Cognition, 10*, 69–75.

Markman, E. M. (1984, October). *How children constrain the possible meanings of words.* Paper presented at a conference on "The Ecological and Intellectual Bases of Categorization," Emory University.

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

McCloskey, M., & Glucksberg, S. (1979). Decision processes in verifying category membership statements: Implications for models of semantic memory. *Cognitive Psychology, 11*, 1–37.

McNamara, T. P., & Sternberg, R. (1983). Mental models of word meaning. *Journal of Verbal Learning and Verbal Behavior, 22*, 449–474.

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

Miller, G. A., & Johnson-Laird, P. N. (1976). *Language and perception.* Cambridge: Harvard University Press.

Minsky, M. A. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision.* New York: McGraw-Hill.

Murphy, G. L. (1987). *Comprehending complex concepts.* Unpublished manuscript, Brown University.

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

Oden, G. G. (1977). Integration of fuzzy logical information. *Journal of Experimental Psychology: Human Perception and Performance. 3*, 565–575.

Oden, G. G. (1981). A fuzzy propositional model of concept structure

and use: A case study in object identification. In G. W. Lasker (Ed.), *Applied systems research and cybernetics* (Vol. 6). New York: Pergamon.

Osherson, D. N., & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition, 11,* 35–58.

Osherson, D. N., & Smith, E. E. (1982). Gradedness and conceptual conjunction. *Cognition, 12,* 299–318.

Putnam, H. (1975). The meaning of "meaning." In K. Gunderson (Ed.), *Language, mind, and knowledge.* Minneapolis: University of Minnesota Press.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization,* (pp. 27–48). Hillsdale, NJ: Erlbaum.

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., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge: Harvard University Press.

Smith, E. E., & Osherson, D. N. (1984). Conceptual combination with prototype concepts. *Cognitive Science, 8,* 337–361.

Smith, E. E., Osherson, D. N., Rips, L. J., Albert, K., & Keane, M. (1985). *Combining prototypes: A modification model.* Unpublished manuscript, Bolt, Beranek, & Newman, Inc., Cambridge, MA.

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.

Thagard, P. (1983, June). *Conceptual combination: A frame-based theory.* Paper presented at the Society for Philosophy and Psychology, Wellesley, MA.

Turksen, I. B. (1984). *Combination of concepts.* Unpublished manuscript, University of Toronto.

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

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293–315.

Wittgenstein, L. (1953). *Philosophical investigations.* New York: MacMillan.

Zadeh, L. (1965). Fuzzy sets. *Information and Control, 8,* 338–353.

Zadeh, L. (1982). A note on prototype theory and fuzzy sets. *Cognition, 12,* 291–297.

Appendix A

Examples From Experiment 1

Item	Response distribution				Mean rating	
	[++]	[+−]	[−+]	[−−]	Conjunction	Category
Office furniture						
Word processor	4	15	0	3	1.64	−1.82
Filing tray	8	14	0	0	2.55	−1.14
Photo-copier	5	14	0	3	2.00	−1.27
Intercom	4	13	0	5	1.05	−1.86
Typewriter	7	11	0	4	2.09	−1.14
Desk blotter	7	11	0	4	2.09	−1.09
Room divider	7	10	1	3	1.14	−0.46
Garden furniture						
Grass roller	1	15	0	6	0.73	−2.36
Bird bath	6	14	0	2	1.82	−1.45
Swing	7	14	0	1	1.95	−1.23
Climbing frame	6	13	0	3	1.45	−1.45
Lawn mower	1	13	0	8	0.82	−2.68
Barbecue	9	12	0	1	1.95	−0.91
Gnome	4	12	0	6	0.91	−1.64
Clothes line	7	12	0	3	1.73	−1.55
Sun umbrella	11	11	0	0	2.27	−0.32
Plant tubs	10	10	0	2	1.86	−0.41
School furniture						
Blackboard	9	13	0	0	2.68	−0.32
Globe	8	12	0	2	1.82	−0.77
Wall chart	3	12	0	7	0.91	−2.14
Bulletin board	9	10	0	3	1.59	−0.82
Church furniture						
Poor box	7	13	0	2	2.14	−1.18
Altar	9	12	0	1	2.73	−0.68
Font	9	12	0	1	2.64	−0.59

Appendix A (*continued*)

Item	Response distribution				Mean rating	
	[++]	[+−]	[−+]	[−−]	Conjunction	Category
Kitchen furniture						
Extractor fan	2	12	0	8	0.55	−2.09
Draining rack	9	12	0	1	2.32	−0.68
Kettle	5	11	1	5	1.14	−1.32
Hot tray	7	11	1	3	0.95	−0.86
Sports vehicles						
Surfboard	6	7	0	9	0.64	−1.32
Polo pony	4	6	1	11	−0.45	−1.36
Wind surfer	11	5	0	6	0.91	−0.59
Aquatic vehicles						
Diving bell	9	6	0	7	1.09	−0.77
Sailing dinghy	15	5	0	2	2.23	0 91
Protective clothing						
Surgeon's mask	8	12	1	1	1.95	−0.73
Cricket pads	9	12	0	1	2.73	−0.68
Fencing mask	10	11	0	1	2.64	−0.50
Rubber gloves	12	10	0	0	2.73	0.23

Note. Plus and minus symbols in brackets refer to conjunction and category judgments, respectively. Only examples with at least 10 [+−] responses are included (except for vehicles, where those with at least five are included).

Appendix B

Overextended Items in Experiment 2

Item	Frequency of inconsistent triple		Probability of being rated		
	Actual	Expected[a]	As sport	As game	In the conjunction
Sport + game − conjunction +					
Javelin	15	4	1.00	.47	.74
Fishing	14	2	1.00	.18	.66
Trampolining	14	4	1.00	.31	.56
Boxing	13	9	1.00	.56	.62
Judo	13	6	1.00	.44	.65
Rowing	13	0	1.00	.29	.71
Fencing	13	6	1.00	.53	.65
Discus	12	8	1.00	.62	.72
Skiing	12	5	1.00	.37	.55
Wrestling	10	9	1.00	.53	.56
Gymnastics	9	8	.94	.50	.47
Archery	8	5	.94	.59	.66
Sport − game + conjunction +					
Bar-billiards	14	0	.35	.94	.79
Chess	12	8	.39	.94	.51
Ten-pin bowling	8	0	.88	1.00	.91
Pool	8	1	.50	.94	.73

[a] On the basis of the stochastic model.

Appendix C

Overextended Items in Experiment 3

Item	Frequency of inconsistent triple		Probability of being rated		
	Actual	Expected[a]	As sport	As game	In the conjunction
Sport + game − conjunction +					
Boxing	23	7	1.00	.43	.66
Fencing	21	1	1.00	.15	.67
Rowing	18	3	1.00	.37	.49
Discus	17	8	.93	.48	.52
Archery	17	6	1.00	.56	.72
Judo	17	2	1.00	.40	.55
Wrestling	17	4	1.00	.45	.54
Trampoline	15	9	.93	.36	.44
Polo	12	1	1.00	.79	.96
Golf	8	1	1.00	.87	.94
Cricket	8	1	1.00	.79	.96
Sport − game + conjunction +					
Dominoes	18	3	.17	1.00	.38
Chess	13	3	.17	1.00	.38
Bridge	12	4	.14	1.00	.28
Pool	11	4	.67	1.00	.79
Hopscotch	10	5	.18	1.00	.31
Hoolahoop	9	10	.48	.96	.37
Noughts & Crosses	8	2	.10	1.00	.20
Bar-billiards	8	7	.68	.96	.69
Ten-pin bowling	8	2	.76	.96	.85
Monopoly	8	6	.14	1.00	.19

[a] On the basis of the stochastic model.

Appendix D

Overextended Items in Experiment 4 for Inconsistent Triple [A+ B− AB+]

Item	Frequency of inconsistent triple		Probability of being rated		
	Actual	Expected	As A	As B	In the conjunction
A. Furniture B. Household appliance					
Filing cabinet	14	2	1.00	.15	.51
Heated water bed	12	2	1.00	.60	.77
Sewing chest	9	5	.95	.45	.50
Coffee table	8	0	1.00	.20	.32
Piano	8	2	1.00	.15	.30
Chair	7	1	1.00	.25	.30
Hifi	6	1	.63	.89	.79
A. Household appliance B. Furniture					
Clothes washer	25	0	1.00	.15	.72
Cooking stove	23	2	1.00	.45	.82
Fridge	18	2	1.00	.45	.77
Vacuum cleaner	15	2	1.00	.05	.37
TV	10	2	.95	.70	.92
Hifi	9	3	.89	.63	.79
Desk lamp	9	4	.90	.75	.82
A. Food B. Plant					
Raisin	15	0	1.00	.40	.77
Mushroom	8	0	1.00	.70	.90
Potato	7	1	1.00	.70	.90
Garlic	6	1	.95	.70	.79
Peppercorn	6	1	1.00	.64	.69
Bread	6	1	1.00	.05	.17
A. Plant B. Food					
Eucalyptus	6	2	.84	.21	.26
A. Weapon B. Tool					
Club	19	5	1.00	.50	.77
Spear	17	1	1.00	.30	.67
Arrow	15	5	1.00	.35	.57
Whip	13	1	.94	.30	.63
Rifle	10	7	1.00	.42	.49
A. Tool B. Weapon					
Screwdriver	15	2	1.00	.30	.62
Hammer	12	3	1.00	.55	.80
Razor	11	2	.85	.65	.82
Chain-saw	10	2	1.00	.55	.75
Chisel	9	1	.95	.45	.62
Scissors	8	0	.95	.50	.79
A. Building B. Dwelling					
Library	8	1	1.00	.20	.35
Theatre	7	1	1.00	.10	.27
A. Dwelling B. Building					
Trailer	15	4	1.00	.35	.60
Tent	9	8	.95	.60	.55
Cave	7	4	.95	.20	.27
Tree house	6	1	.95	.80	.85
A. Machine B. Vehicle					
Ski lift	11	0	1.00	.70	.87
Elevator	10	2	.95	.75	.80
Lawn-mower	7	0	1.00	.10	.26